Early genomic response to X-ray induced DNA damage and transposon regulation in *Arabidopsis thaliana*

Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie der Freien Universität Berlin

vorgelegt von

Zhenxing Wang

August 2014

Die vorliegende Arbeit wurde in der Zeit von Oktober 2009 bis August 2014 unter der Leitung von Prof. Dr. Reinhard Kunze am Institut für Biologie - Angewandte Genetik angefertigt.

1. Gutachter: Prof. Dr. Reinhard Kunze
   Freie Universität Berlin
   Institut für Biologie – Angewandte Genetik

2. Gutachter: PD. Dr. Alexander Heyl
   Freie Universität Berlin
   Institut für Biologie – Angewandte Genetik

   Disputation am 20. November 2014

Acknowledgements

I am grateful to my supervisor Prof. Dr. Reinhard Kunze for giving me a chance to study in Germany. Thanks for his ideas, support, and suggestions which inspired and helped me to finish my work throughout the years. I also want to thank PD. Dr. Alexander Heyl to be my second supervisor. Thank my supervisors taught me the rigorous german attitude in scientific work.

I want to thank all my friendly colleagues. They helped me to get used to work in the lab quickly, to deal with the troubles in experiments. Because of them, I enjoyed the working atmosphere in the institute. In particular, I thank Dr. Anahid Powell for her kind advices and all favors of my work.

I want to thank Dr. Evelyn Klocke and Ms. Weinzierl at the Julius Kühn-Institut (JKI) for kindly supplying radiative machine and helping me to treat seedlings with X-ray. I thank Dr. Sebastian Proost and Dr. Han Wang for helping me to set up the working environment of RNA-seq on Linux background. I want to thank Dr. Elke Diederichsen for discussing and improving statistic analysis.

I want to thank China Scholarship Council, Dahlem Centre of Plant Sciences and Freie Universität Berlin for financial support.

I want to thank Mr. Richard Hart-Jones, Dr. Alexis Kasaras, Dr. Christine Rausch, Dr. Diana Mutz, Dr. Jiang Zhang, Dr. Ling Chang, Kevin Mielich, Navina Drechsler, Jana Oberländer and Yue Zheng for correcting and improving my writing in thesis.

I want to thank all my friends for their support.

Finally, I am grateful to my mother and all my family members. Without their long lasting support and encourage, I would not be able to finish this work. I wish my father could see my achievements in heaven as well.

致谢

五年前的一日，湘中小城邵阳尚是炎夏。跨越七千公里，到达大陆另一端的异国首都，柏林却已然洋溢着北地初秋的清爽。蔚蓝的天空，悠悠地缀着些白云，留学德意志的日子就在这样一个平凡得不能再平凡的秋日开始了。未曾料到的是，五年的时光转瞬即逝，一切都只是为了追求人类目前最高的学位。

坚硬的德国面包，散发奇特香味的奶酪，勾起对家乡牛肉粉的无比想念，当然那更是对家人的思念。感谢坚强的母亲在父亲逝去后，仍支持我远赴他国读博。愧疚于母亲有病痛时只能自己默默承受，身为儿子却不能随侍身旁。也要感谢所有家人对我一直在外求学的支持。在这漫长的求学历程里，父亲，爷爷，外公，奶奶相继驾鹤归去，希望这篇博士毕业论文能使他们得到告慰。

感谢时空相隔，在地球另一侧也关心和鼓励我奋斗的同学和朋友们。在此也要感谢在柏林结识的朋友们。我们彼此扶持为学业而奋斗；为欣赏欧罗巴各国的风情而组团穷游；共同体味异国生活的酸甜苦辣。这必是我生命中将永久铭记的美好岁月。

Contents I
List of Figures IV
List of Tables VI
Abbreviations VII

Contents

1 Introduction 1
1.1 Transposable Elements (TEs) 1
1.1.1 Functions of TEs in genome 3
1.1.2 Silencing of transposons 4
1.1.3 Silent transposable elements can be reactivated 10
1.2 DNA damage repair is associated with epigenetic modifications 11
1.2.1 Origins of DNA damage and repair pathways 12
1.2.2 Epigenetics in single-strand break repair 13
1.2.3 Epigenetics in double-strand breaks repair 14
1.3 Chromatin remodeling 16
1.4 Mis-processing of *Ac* transcripts in *Arabidopsis* 17
1.5 RNA-seq technology 19
1.6 Goals of the research 22
2 Materials and Methods 23
2.1 Materials 23
2.1.1 Antibodies, Enzymes and Kits 23
2.1.2 Ladders 24
2.1.3 Oligonucleotides 24
2.1.4 Plasmids 25
2.1.5 Bacteria strains 26
2.1.6 Plant material 27
2.1.7 Gene synthesis 28

2.1.8 Sequencing 28
2.1.9 Computer programs and database 28
2.2 Methods 29
2.2.1 Handling bacteria and plants 29
2.2.2 Molecular biological methods 31
2.2.3 X-ray irradiation treatment 38
2.2.4 RNA-seq and bioinformatics analysis 39
3 Results 44
3.1 Generating DSBs by meganucleases 44
3.1.1 Generation of *Arabidopsis* DSB reporter plants 44
3.1.2 Estradiol induced expression of ISI-opA generated DNA damage 48
3.1.3 Meganuclease induced DSBs did not result in transcriptional changes of detected genes 52
3.2 Ionizing radiation induced transcriptomic response 55
3.2.1 X-ray irradiation of plants 56
3.2.2 Overview of RNA-seq data 56
3.2.3 Overview of regulated transcripts in response to X-ray 59
3.2.4 X-ray induced reactivation and suppression of TEs 67
3.2.5 Novel transcripts are all long non-coding RNAs 69
3.2.6 ATM regulates lncRNAs in response to DSBs 72
3.2.7 Co-regulation of lncRNA-flanking genes 73
3.2.8 DNA methylation of selected loci and immunodetection of Histone modifications after X-ray irradiation 76
3.2.9 AGOs in response to X-ray induced DNA damage 78
3.3 Further analysis of DRM1, JMJ30 and AGOs in response to DSBs 81
3.3.1 Genotyping of selected mutants 81
3.3.2 Mutants responded differently to X-ray irradiation 83
3.3.3 DRM1 and BRCA1 affect TEs and lncRNAs in response to DSBs 84
3.3.4 Regulation of lncRNAs and TEs in *JMJ30*, *AGO2*, and *AGO4* defective plants 86
3.4 Regulation of heterologous TE 89

3.4.1 Generating *Ac* mutants and transgenic plants 89
3.4.2 The *miniDs* excision assay 90
4 Discussion 95
4.1 Meganuclease induced DSBs 95
4.1.1 Estradiol induced expression of meganuclease could control the generation of DSBs in *Arabidopsis* 95
4.1.2 Meganuclease-induced DSB frequency is insufficient for detecting alterations in transposon activity or epigenetic status 96
4.2 Early responses to X-ray induced DNA damage in *Arabidopsis* 98
4.2.1 RNA-seq reveals a complete overview of early transcriptomic responses to DSBs 99
4.2.2 TEs are insensitive to DSBs 100
4.2.3 LncRNAs are associated with DSBs response 102
4.2.4 TEs/repeats-derived lncRNAs affect associated and nearby genes 103
4.2.5 *DRM1* might be involved in DSBR and regulates lncRNAs downstream of *BRCA1* 104
4.2.6 Histone modifications were not generally altered 3 hours after X-ray irradiation 106
4.2.7 Does *JMJ30* participate in DSB-induced RdDM? 108
4.2.8 Multiple functions of AGOs in response to DSBs? 109
4.2.9 MicroRNA genes 110
4.2.10 Model of TEs/repeats in response to DSBs 111
4.3 *Ac*, an example of defending heterologous TEs in *Arabidopsis*? 112
5 Summary 115
6 Zusammenfassung 117
7 References 119
8 Appendix 138

List of Figures

Figure 1	Types of transposable elements and mechanism of transposition	2
Figure 2	Transposons in *A. thaliana* (percentage of total numbers)	3
Figure 3	Methylation of cytosine	5
Figure 4	Epigenetic mechanisms in plants	7
Figure 5	Silent TEs can be reactivated by “genomic stress”	11
Figure 6	Positions of *Ac* mis-processed transcripts detected in *Arabidopsis* and sugar beet	18
Figure 7	A typical RNA-seq experiment	20
Figure 8	Structure of binary vectors and various target sites interrupted GUS CDSs	45
Figure 9	Southern blotting analysis of transgenic plants	47
Figure 10	Detecting transcripts of nucleases after different induced periods	48
Figure 11	Illustration of estradiol induced DSBs in plants	49
Figure 12	GUS staining of estradiol induced plants	50
Figure 13	*In vitro* analysis of DSBR	51
Figure 14	Repairing patterns of *ISI-opA* encoded I-Sce I induced damage in plants	52
Figure 15	qRT-PCR detection of selected targets in estradiol induced ISI-opA plants	55
Figure 16	Plants for X-ray treatment and X-ray machine	57
Figure 17	Density plot of individual conditions	58
Figure 18	Venn diagrams of regulated transcripts by the induction of X-ray irradiation	59
Figure 19	Partial hieratical graph of enriched Gene Ontology (GO) terms of upregulated genes in WT	62
Figure 20	Partial hieratical graph of enriched GO terms of upregulated genes in *atm*	63
Figure 21	Partial hieratical graph of enriched GO terms of downregualted genes in WT	63
Figure 22	Partial hieratical graph of enriched GO terms of downregualted genes in *atm*	64
Figure 23	Stress associated genes responded to DSBs in WT (A) and *atm* mutant (B)	65
Figure 24	DSBs induced active DNA and histone methylation etc.	66
Figure 25	Regulated transposons in WT plants	68
Figure 26	qRT-PCR detection of TEs/repeats	69
Figure 27	Visualizing RNA-seq data of RLFS_026432 and At4NC060340 by IGB	71
Figure 28	Detected expressions of lncRNAs RLFS_026432 and At4NC060340 by qRT-PCR	71
Figure 29	Venn diagram of regulated lncRNAs in WT and *atm* mutant	73
Figure 30	Illustration of AT1TE69975 and AT1TE69990 located in introns of *WRR4* gene and semi RT-PCR confirmation	74
Figure 31	Illustration of AT1TE04710 and AT1TE04720 location and visualization of reads mapped to this region	75
Figure 32	qRT-PCR detection of *MET1*, *CMT3*, *DRM1*, *DRM2*, *JMJ30*, *ASHR3*, and *ATXR6*	77
Figure 33	Cytosine methylation of AT1TE69975, AT1TE69990 and promoter region of *WRR4*	79
Figure 34	Histone modifications in WT and *atm* mutant 3 hours post X-ray treatment	80
Figure 35	qRT-PCR detection of *AGO2*, *AGO4* and *AGO7*	80
Figure 36	Gene structures and T-DNA insertions	82
Figure 37	Genotyping PCR of T-DNA mutant lines	82
Figure 38	Phenotypes of mutant plants post irradiation	84

Figure 39 Transcripts of TEs and lncRNAs detected in *brca1* and *drm1* mutants by qRT-PCR 86
Figure 40 Detected transcriptional levels of TEs and lncRNAs in *jmj30*, *ago2*, and *ago4* mutants 88
Figure 41 Construction of *Ac* cDNA derivatives with eliminated mis-processing sites 91
Figure 42 Southern blots of *AcTPase* WT and derivatives plants 92
Figure 43 *Ds* excision screening 93
Figure 44 Detected *Ac* transcripts in selected plants by RT-PCR 94
Figure 45 Hypothetical pathway of BRCA1 regulating MET1 and DRM1 in lncRNAs biosynthesis 105
Figure 46 Schematic illustration of potential responses of TEs/repeats to DSBs 111

List of Tables

Table 1	Proportions of transposable elements in genomes of selected organisms	2
Table 2	Examples of epigenetic regulated transposable elements	9
Table 3	DNA damage related epigenetic events	15
Table 4	Antibodies applied	23
Table 5	Kits applied	24
Table 6	Tables of primers listed in appendix	24
Table 7	Plasmids list	25
Table 8	List of generated transgenic plants	27
Table 9	Commonly used computer programs	28
Table 10	List of programs for RNA-seq analysis	28
Table 11	Online tools	29
Table 12	Ratios of mapped reads	58
Table 13	Non-coding transcripts predicted by CPC	70
Table 14	Differentially expressed novel transcripts in WT and *atm* plants	70
Table 15	Ds transposition efficiencies in transgenic plants	93
Table S1	Primers for generating constructs of estradiol induced cleavage system	138
Table S2	Primers for *Ac*	138
Table S3	Primers for qRT-PCR	139
Table S4	Primers for T-DNA insertion lines	141
Table S5	Primers for Bisulfite sequencing analysis	141
Table S6	Go enrichment comparison of upregulated genes in WT and *atm* mutant	141
Table S7	Go enrichment comparison of downregulated genes in WT and *atm* mutant	143
Table S8	Regulated TEs/repeat elements in WT	147
Table S9	Regulated TEs/repeat elements in *atm*	150
Table S10	LncRNAs regulated by X-ray irradiation and co-expressed genes in WT	151
Table S11	LncRNAs regulated by X-ray irradiation and co-expressed genes in *atm* mutant	156

Abbreviations

5-Aza-CdR	5-aza-2‘-deoxycytidine
A. thaliana	*Arabidopsis thaliana*
Ac	Activator, autonomous transposable element from Maize
AcTPase	Ac transposase
AG	AGAMOUS
AGO	ARGONAUTE
AP2	APETALA2
ASH	absent, small, or homeotic discs
ASHH1	ASH1-HOMOLOG 1
ASHR3	ASH1-RELATED 3
ATM	Ataxia telangiectasia mutated
ATX	*Arabidopsis* trithorax
BER	excision repair
BRCA1	breast cancer susceptibility 1
C. elegans	*Caenorhabditis elegans*
CaMV	cauliflower mosaic virus
CBP	CREB-binding protein
CLF	CURLY LEAF
CMT3	chromomethyltransferse 3
CPC	Coding Potential Calculator
CPD	cyclobutane pyrimidine dimer
CtIP	C-terminal binding protein interacting protein
DCL3	DICER-LIKE 3
DDM1	DECREASED DNA METHYLATION 1
DDR	DNA damage repair
diRNA	DSB-induced small RNA
DME	DEMETER
DML	DEMETER-LIKE
DIG	digoxygenin
DNA-PK	catalytic subunit of a nuclear DNA-dependent serine/threonine protein kinase
DRM	Domains Rearranged Methyltransferase
D.melanogaser	*Drosophila melanogaster*
Ds	Dissociation, non-autonomous transposable element from Maize
DSB	double-strand break
DSBR	double-strand breaks repair
dsRNA	double-stranded RNA
EMSA	electrophoretic mobility shift assay
FPKM	fragments per kilobase of exon sequences per million mapped reads
FWA	FLOWERING WAGENINGEN
GATU	gene-associated transcription unit
GO	Gene Ontology

H. sapiens	*Homo sapiens*
H3K9	histone H3 lysine 9
H3K9me2	di-methylation at histone H3 lysine 9
HAM	histone acetyltransferases of the MYST family
hAT	Abrreviation of hobo, Activator, and Tam3
HAT	histone acetyltransferase
HDAC	histone deacetylase
HDM	Histone demethylase
HKMT	Histone methyltransferase
HMT	histone methyltransferase
HR	homologous recombination
HRP	horse radish peroxidase
ING1	Inhibitor of growth 1
IR	Ionizing radiation
I-Sce I	homo endonuclease derived from yeast
ISI-opA	optimized CDS for I-Sce I protein in *Arabidopsis*
JMJ	Jumonji C domain-containing protein
KDM5B	Lysine-specific histone demethylase 5B
KYP	KRYPTONITE
LEDGF	Lens epithelium - derived growth factor
LINE	long interspersed element
lncRNA	long non-coding RNA
LSD	Lysine Specific Demethylase
LSD1	Lysine Specific Demethylase 1
LTR	long terminal repeat
M3	Ac CDS/transgenic plant contains 3 point mutations
M7	Ac CDS/transgenic plant contains 7 point mutations
MBD	methylated DNA-binding protein
MEA	MEDEA
MES	2-(N-morpholino) ethanesulfonic acid
MET1	Methyltransferase 1
miRNA	microRNA
MMEJ	microhomology-mediated end joining
MMR	mismatch repair
MMS	methylmethane sulfonate
MRN	MRE11/RAD50/NBS1 complex
MULE	Mutator-like element
NAT	natural antisense transcript
ncRNA	non-coding RNA
NER	nucleotide excision repair
NHEJ	non-homologous end joining
O. sativa	*Oryza sativa*
PBA	4-phenylbutyric acid
PLncDB	plant long non-coding RNA database

Pol	RNA polymerase
PTGS	post-transcriptional gene silencing
QQR	zinc finger nuclease
qRT-PCR	Quantitative RT-PCR
RCTU	repeat-containing transcription unit
RdDM	RNA-directed DNA methylation
RDR2	RNA-dependent RNA polymerase 2
REF6	RELATIVE OF EARLY FLOWERING 6
ROS	Reactive oxygen species
ROS1	repressor of silencing 1
S. cerevisiae	*Saccharomyces cerevisiae*
SDG	SET Domain Group
SDSA	synthesis-dependent strand-annealing
SET	Suppressor of variegation (Su(var)3-9),Enhancer of Zeste (E(z)) and Trithorax (TRX)
SINE	short interspersed element
siRNA	small interfering RNA
SOG1	Suppressor of gamma response 1
sRNA	small RNA
SSB	single-strand break
ssRNA	single-strand RNA
SUP	SUPERMAN
SUVH	SU(VAR)3-9 homolog
SWN	SWINGER
ta-siRNA	trans-acting siRNA
TDG	thymine DNA glycosylase
TE	transposable element
TGS	transcriptional gene silencing
TIR	terminal inverted repeat
TOE3	TARGET OF EAT 3
uidA	² -Glucuronidase coding gene
WRR4	White Rust Resistance 4
Z. mays	*Zea mays*
ZFN3	zinc finger nuclease 3

1 Introduction

1.1 Transposable Elements (TEs)

Transposable elements are DNA fragments which can move from their orginal locations and (re)integrate into new postitions in genome via either reverse transcription process or “cut-and-paste” mechanism. Therefore they can be classified into two groups: 1) Class I TEs are retrotransposons that replicate by a reverse-transcription process and integrate cDNA copies of the original element in the genome. This group includes *gypsy*-like retrotransposons and *copia*-like retrotransposons with long terminal repeats (LTRs), long interspersed elements (LINEs) and short interspersed elements (SINEs), which are Non-LTR retrotransposons. 2) Class II TEs are DNA transposons. Most, but not all class II TEs are mobilized by a “cut-and-paste” mechanism. DNA transposons are categorized into many different families. The most prominent ones are the *Mutator*-like element (*MULE*), *hAT* family, *CACTA*-like transposon, *Tc1*/*mariner* family, and *P elements*. A distinct subgroup is the Helitron, which transpose via a rolling-circle mechanism (Kapitonov and Jurka 2001), but not a cut-and-paste mechanism. Principal structures of TEs and mechanism of transposition are illustrated in Figure 1 (Lisch 2013).

TEs were first discovered in maize by McClintock in the late 1940s (McClintock 1949, McClintock 1950). In subsequent decades, TEs were found to spread widely in organisms, however, their proportion in different eukaryotic genomes varies from 2 % in the pufferfish (Elgar et al. 1999) to 85 % in maize (Table 1) (Baucom et al. 2009, Schnable et al. 2009). *Arabidopsis thaliana* has a small genome with a rather low TE content (~17 %), but it contains all the typical families of transposable element (Buisine et al. 2008) (Figure 2).

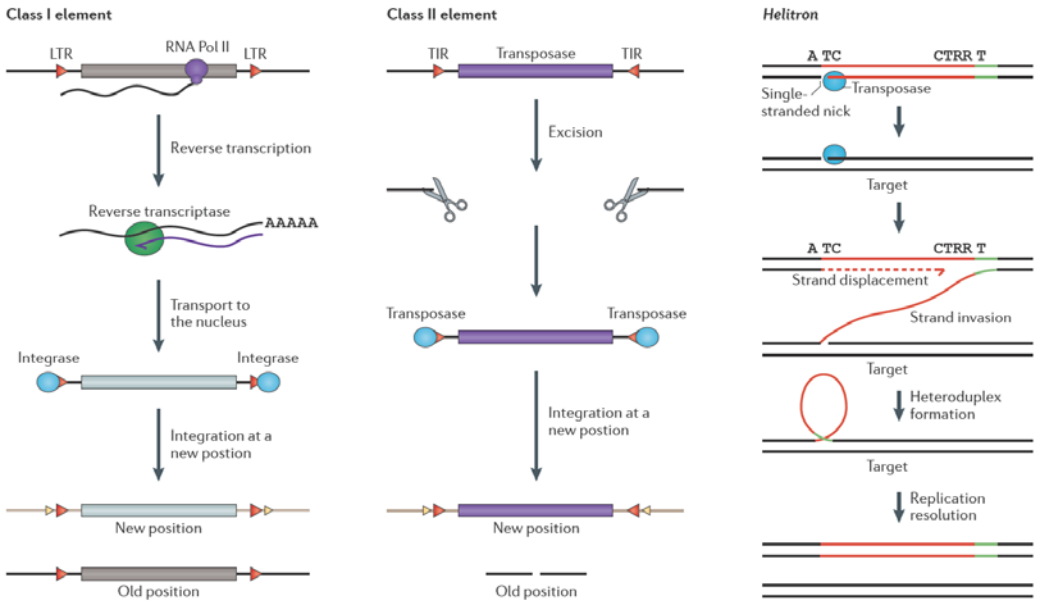


Figure 1: Types of transposable elements and mechanism of transposition (Lisch 2013).

Class I elements: retrotransposons. The Retrotransposon is reverse transcribed to a cDNA and then inserts into a new position in the genome by mediation of integrase. LTR: long terminal repeat. Integrase: enzyme facilitating integration. Class II elements: DNA transposons. These elements excise from their original position at the presence of transposase and further reinsert into new positions in the genome. TIR: terminal inverted repeat. Transposase: coding sequence (CDS) of transposase. Helitron: DNA transposon mobiles via rolling-circle mechanism.

Table 1: Proportions of transposable elements in genomes of selected organisms

Species	%TEs of genome	References
Arabidopsis thaliana	17 %	(Buisine et al. 2008)
Zea mays	85 %	(Baucom et al. 2009, Schnable et al. 2009)
Hordeum vulgare L	>70 %	(Park et al. 2004)
Oryza sativa	20 %	(Turcotte et al. 2001)
Mus musculus	38 %	(Waterston et al. 2002)
Homo Sapiens	50 %	(Lander et al. 2001)
Fugu rubripes	2 %	(Elgar et al. 1999)
Danio rerio	27 %	(Sela et al. 2010)
Tribolium castaneum	33 %	(Richards et al. 2008)
Aedes aegypti	47 %	(Nene et al. 2007)
Bombyx mori	35 %	(Osanai-Futahashi et al. 2008)
Drosophila melanogaster	15 %	(Adams et al. 2000)
Saccharomyces cerevisiae	3 %	(Carr et al. 2012)

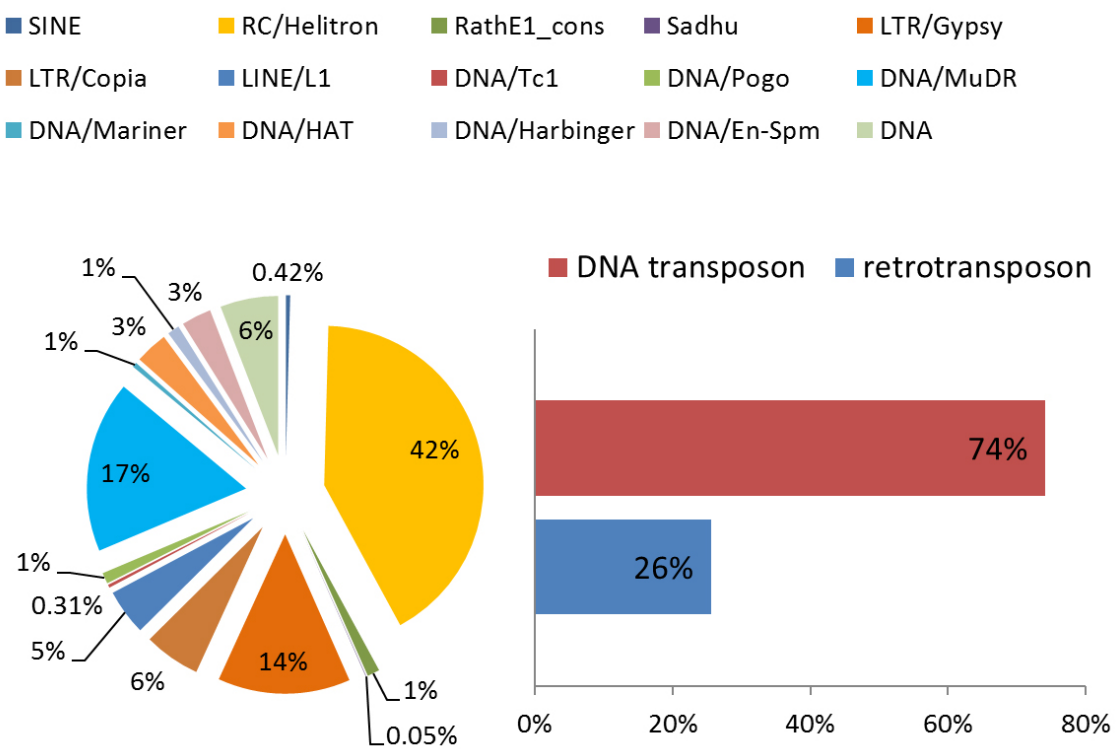


Figure 2: Transposons in *A. thaliana* (percentage of total TE numbers)

The charts are generated based on data collected from The *Arabidopsis* Information Resource (TAIR) (Lamesch et al. 2012). *Copia*-like (6 %) and *Gypsy*-like (14 %) are retrotransposons with long terminal repeats (LTRs). Short interspersed element (SINE) (0.42 %), long interspersed element (LINE) (5 %), RathE3_cons (1 %), and Sadhu (0.05 %) belong to non-LTR retrotransposons. *Tc1* (0.31), *Pogo* (1 %), *MuDR* (17 %), *Mariner* (1 %), *HAT* (3 %), *Harbinger* (1 %), *En-Spm* (3 %), and RC/Helitron (42 %) belong to DNA transposon. The remaining DNA transposable elements account for 6 %. In general, percentages of retrotransposon and DNA transposon are 25.57 % and 74.12 %, respectively. The remaining 0.33 % is unsigned element. Percentage represents proportion of each TE family of all predicted TEs in *A. thaliana*.

1.1.1 Functions of TEs in genome

Despite of existing widely in genome of different species, TEs were labeled as junk DNA for decades (Orgel and Crick 1980). However, accumulating evidences indicate that TEs play roles in regulating gene expression and genome evolution.

Many TEs contain gene fragments or inserted close to or even in genes, thus act as regulatory elements. In Maize, *Doppia* transposon separates duplications of *R-r* allele and leads to anthocyanin pigment expression in the seed (Walker et al. 1995). Insertion of *Gret1 LTR* element upstream of *Vvmbya1* gene inhibits grape colour (Kobayashi et al.

2004). Light responsive transcription factors genes, *FAR1* and *FHY3*, were reported to be regulated by domesticated TEs (Hudson et al. 2003). Gene expression is regulated by epigenetic manipulation, in which small RNA mediated silencing is one of the pathways. In *Arabidopsis*, a large number of RNA polymerase V (Pol V) generated small RNAs are derived from SINE repeats (Lee et al. 2012). TEs originated small RNAs could reinforce TEs silencing at other loci or affect gene expression via targeting homologous sequences in regulatory region (Feng et al. 2010).

Excision and reinsertion may lead to genome instability. On the other hand, the DNA breakage facilitates the rearrangement of chromosomes which can drive evolution of genome. TEs are considered to affect genome size, gene structure, and create new genes through their mobile capabilities (Bennetzen and Wang 2014).

1.1.2 Silencing of transposons

Transposons constitute a large proportion of the genomic DNA, but most of them are quiescent under normal conditions. The maintenance of the silenced state of TEs is considered to be important in order to prevent the accumulation of novel and potentially deleterious mutations, which are caused by excision and reinsertion of transposons in genes. Quiescent TEs are both transpositionally and transcriptionally inactive. In order to prevent the destruction of genome integrity by frequent transposon insertion mutations, a set of sophisticated epigenetic mechanisms is recruited to silence transposable elements.

1.1.2.1 Epigenetic mechanism

Epigenetic regulation, including DNA methylation and histone modification, occurs genome-widely. In both, mammals and plants, cytosine can be methylated to 5m-cytosine (Figure 3) (Bird et al. 1995) in three different contexts CG, CHG and CHH (H represent A, C, or T). Symmetrical CG methylation is maintained by Methyltransferase 1 (MET1) (Kankel et al. 2003) and can be transmitted stably over

generations (Schmitz et al. 2011). MET1 acts behind the DNA replication fork on newly replicated, hemi-methylated DNA as a substrate (Bucher et al. 2012). CHG methylation is established by the activity of *chromomethyltransferse 3* (*CMT3*) with the cooperation of *KRYPTONITE* (*KYP*) which catalyzes the dimethylation of histone H3 lysine 9 (H3K9). It has been described that a defect in *CMT3* leads to decreased CHG methylation at several loci in *Arabidopsis* (Johnson et al. 2002, Jackson et al. 2004). Like CG methylation, CHG methylation can be maintained after DNA replication (Bucher et al. 2012). The formation of asymmetrical CHH methylation is accomplished by the concerted action of the RNA-directed DNA methylation (RdDM) pathway and the *Domains Rearranged Methyltransferase 2* (*DRM2*) (Matzke and Birchler 2005). The reversible reaction, active DNA demethylation is executed by DEMETER-LIKE DNA glycosylase family, which includes *repressor of silencing 1* (*ROS1)*, *DEMETER* (*DME)*, *DEMETER-LIKE 2* (*DML2)* and *DEMETER-LIKE 3* (*DML3)* (Law and Jacobsen 2010) (Figure 4).

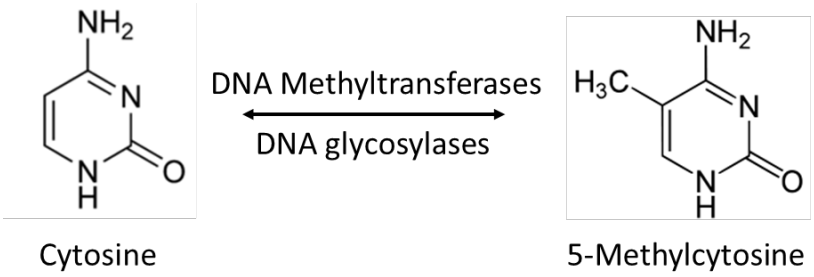


Figure 3: Methylation of cytosine.

Methylation is a process that adds an additional methyl group to a substrate or the replacement of an atom or group by a methyl group. The 5th carbon of cytosine could be the destination for methylation by DNA methyltransferases.

Acetylation and methylation of histones are two remarkable epigenetic modifications that regulate gene expression (Figure 4). Acetylation is introduced to histone by acetyltransferases (HATs). Its reversible process is mediated by histone deacetylases (HDACs) (Ma et al. 2013b). HATs include GNAT, MYST, CBP/p300, and TAF1/TAFII250 families (Berr et al. 2011). They mediate acetylation at lysine 144 of histone H2A (H2AK144ac), lysine 6, 11, 27 and 32 of H2B (H2BK6ac, H2BK11ac, H2BK27ac and

H2BK32ac), lysine 9, 14, 18, 27 and 56 of H3 (H3K9ac, H3K14ac, H3K18ac, H3K27ac and H3K56ac), and at lysine 5, 8, 12, 16 and 20 of H4 (H4K5ac, H4K8ac, H4K12ac, H4K16ac and H4K20ac) (Earley et al. 2007, Zhang et al. 2007). HDACs can be classified into RPD3/HDA1, HD2-like, and SIR2 families (Berr et al. 2011, Ma et al. 2013b).

The SET domain (Suppressor of variegation (Su(var)3-9),Enhancer of Zeste (E(z)) and Trithorax (TRX)) was first identified and named in *Drosophila* histone methyltransferases (HMTs). The SET domain containing HMTs catalyze specific lysine residues of histone to be mono-, di-, and tri-methylated (Liu et al. 2010a, Berr et al. 2011). In the *Arabidopsis* genome 47 SET Domain Group (SDG) proteins are predicted (Ng et al. 2007). They have specific target lysine residues, for instance, *Arabidopsis* trithorax 1 (ATX1), ATX2 and SDG4 are responsible for H3K4 methylation. KYP, SU(VAR)3-9 homolog 5 (SUVH)5, SUVH6, *Arabidopsis* trithorax related protein 4 (SUVR4) are in charge of H3K9 methylation. ATXR5, ATXR6, CURLY LEAF (CLF), MEDEA (MEA), SWINGER (SWN) methylate H3K27. H3K36 methylation is catalyzed by SDG4, SDG8 and SDG26 (Liu et al. 2010a). The erasers of histone methylation are Lysine Specific Demethylase (LSD) proteins and Jumonji C domain-containing protein (JMJ). They have been reviewed recently (Chen et al. 2011).

Various histone acetylation and methylation are hallmarks of gene expression status. H3K9ac, H3K27ac, H3K4me3, and H3K36me2/me3 are histone marks associated with active transcription. H3K9me2 and H3K27me3 are marks of suppressed gene transcription (Lauria and Rossi 2011).

Besides acetylation and methylation, ubiquitination of histone was reported to promote DNA methylation and heterochromatin formation as well (Sridhar et al. 2007).

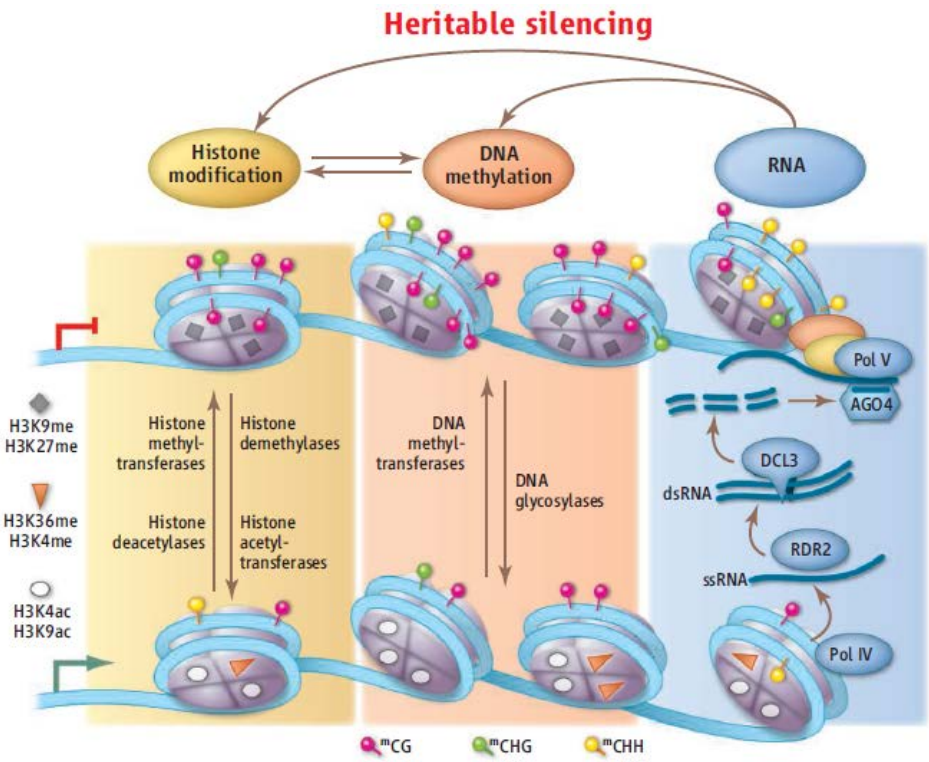


Figure 4: Epigenetic mechanisms in plants (Fedoroff 2012)

There are three cytosine methylation contexts of DNA sequences: CG, CHG and CHH (H: A, T or C). DNA methylation and demethylation is a reversible process which is catalyzed by DNA methyltransferases and glycosylases. DNA methylation is correlated with histone modifications, for instance, acetylation at lysine residues 4 and of histone H3, methylation at lysine residues 4, 9, 27, and 36 of histone H3. Histone acetyltransferases, deacetylases, methyltransferases, and demethylases are the functional proteins for these reactions. Plant specific RNA polymerase IV (Pol IV) and Pol V participate in RNA-directed DNA methylation (RdDM) in cooperation with Argonaute 4 (AGO4), AGO6, DICER-LIKE 3 (DCL3), and RNA-dependent RNA polymerase 2 (RDR2).

Small RNA is reported to participate in the process of histone modification and DNA methylation. Plant-specific RNA polymerase IV (Pol IV) generated single-strand RNA (ssRNA) is loaded by RNA-dependent RNA polymerase 2 (RDR2) to form double-stranded RNA (dsRNA), and then cleaved into short 21-26-nucleotide small interfering RNAs (siRNAs) by DICER-LIKE 3 (DCL3). Argonaute 4 (AGO4) and another plant-specific RNA polymerase, Pol V, recruit siRNAs and introduce DNA methylation at target loci via DRM2 (Fedoroff 2012) (Figure 4). A recent article mentioned that the maintenance of DRM2 mediated RdDM is regulated by an H3K4me3 demethylase, JMJ14 (Deleris et al. 2010). In mammals, small RNA mediates gene silencing through

Piwi-piRNA complexes (He et al. 2011, Siomi et al. 2011).

Epigenetic mechanism controls gene expression in animals and plants. It contributes to the regulation of TEs transcription and transposition as well.

1.1.2.2 Epigenetic regulation of TEs

Most transposons are epigenetically silenced, in order to prevent the damage caused by excision and reinsertion of active TEs. It is a common mechanism that regulates TEs in creatures. Table 2 collects epigenetic events which are associated with the alteration of activities of transposable elements.

DNA methylation pattern of *Ac/Ds* was first described almost two decades ago (Wang et al. 1996). Subsequent research demonstrated that decreased DNA methylation promotes *Ds* excision in petunia protoplasts (Ros and Kunze 2001). The status of promoter methylation is associated with activity of *Spm* (Banks et al. 1988). This process is regulated by *Spm*-encoded TnpA protein (Cui and Fedoroff 2002). The DNA methylation pattern of *MuDR* was analyzed in maize as well (Lisch et al. 2002). The status of *Tos17* is regulated by DNA methylation and H3K9me2 in rice (Ding et al. 2007, La et al. 2011). JMJ703, a rice H3K4 demethylase, controls activity of *LINE1*, but not *Tos17* (Cui et al. 2013). In *Arabidopsis*, Lippman and colleagues observed that transcription of *AtMu1*, *AtCOPIA4*, *AtLINE1-4*, *AtLANTYS*, and *AtGP1* were regulated by decreased DNA methylation 1 (DDM1), MET1 and HDA6 (histone deacetylase 6, also named sil1), and the authors speculated that MET1 might interact with HDA6 (Lippman et al. 2003). Eight years later, the cooperation of MET1 and HDA6 and in turn, inhibited transposon activity was finally proven by another group (Liu et al. 2012). The regulation of other TEs by DNA methylation was reported as well (Miura et al. 2001, Kato et al. 2003, Tsukahara et al. 2009). Reports also revealed that RdDM machinery components, DRD1, Pol IVb, and RDM4 controlled *IG/LINE* and *ATCOPIA95* (AT2G04460) (Huettel et al. 2006, He et al. 2009). ARGONAUTE 4 (AGO4), AGO6, and AGO9, are associated with endogenous 24 nt sRNAs in order to silence loci containing TEs, repetitive DNA elements, and heterochromatin regions with partial redundancy

(Slotkin and Martienssen 2007).

In *C. elegans*, RNAi induces silencing of the *Tc1* element (Sijen and Plasterk 2003). In *D.melanogaster*, PIWI family members, PIWI, Aubergine, and AGO3, regulate activities of LTR-retrotransposons and non-LTR elements (Vagin et al. 2004, Kalmykova et al. 2005, Li et al. 2009a). *Alu* elements are well known human transposons, they are heavily methylated at cytosine, and accompanied with dimethylation at H3K9 (Kondo and Issa 2003, Xie et al. 2009).

Table 2: Examples of epigenetic regulated transposable elements

Name	Species	Conditions	Controlling Genes	Reference
Ac/Ds	*Z. mays*	DNA methylation	-	(Wang et al. 1996)
MuDR	*Z. mays*	DNA methylation	*MOP1*	(Lisch et al. 2002)
Tos17	*O. sativa*	DNA demethylation, H3K9me2 methylation	*DNG701, SDG714*	(Ding et al. 2007, La et al. 2011)
LINE1	*O. sativa*	H3K4me3 demethylation	*JMJ703*	(Cui et al. 2013)
AtMu1, AtCOPIA4	*A. thaliana*	DNA methylation; Histone deacetylation	*DDM1,MET1, CMT3, KYP; HDA6*	(Lippman et al. 2003)
AtLINE1-4, AtLANTYS, AtGP1	*A. thaliana*	DNA methylation; Histone deacetylation	*DDM1, MET1; HDA6*	(Lippman et al. 2003)
AtCOPIA13, AtCOPIA21, AtCOPIA93, AtGP3	*A. thaliana*	DNA methylation	*DDM1*	(Tsukahara et al. 2009)
CACTA	*A. thaliana*	DNA methylation	*DDM1, MET1, CMT3*	(Miura et al. 2001, Kato et al. 2003)
IG/LINE	*A. thaliana*	RdDM	*DRD1, Pol IVb*	(Huettel et al. 2006)
ATCOPIA95 (AT2G04460)	*A. thaliana*	RdDM	*RDM4*	(He et al. 2009)
HARBINGER, ATENSPM2, ATCOPIA72, AT4G09480,	*A. thaliana*	Histone deacetylation, DNA methylation	*HDA6, MET1*	(Liu et al. 2012)

AT2G04460				
Tc1	*C. elegans*	RNAi	*MUT-7, MUT-16, pk732*	(Sijen and Plasterk 2003)
copia, HeT-A and I elements	*D.melanogaster*	RNAi	*Aubergine*	(Vagin et al. 2004)
gypsy, ZAM, and idefix	*D.melanogaster*	RNAi	*AGO3*	(Li et al. 2009a)
mdg1	*D.melanogaster*	RNAi	*PIWI*	(Kalmykova et al. 2005)
Alu,	*H. sapiens*	DNA methylation, H3K9me3	-	(Kondo and Issa 2003, Xie et al. 2009)

Cytosines can be methylated not only in TEs/repeats, but also in normal genes. In animals, there are no homologs of plant CMT3 and DRM2 proteins. Therefore, in animals CG methylation dominates both in genes and TEs/repeats. Also in plants, the gene-body methylated cytosines occur only in CG sequence, whereas in TEs/repeats sequences are found in CG, CHG, and CHH sequences (Feng et al. 2010, Zemach et al. 2010, Saze and Kakutani 2011)

1.1.3 Silent transposable elements can be reactivated

Generally, transposons in a genome are silenced globally by the epigenetic machinery. However, reversion of their activities under biotic and abiotic stresses (Figure 5) was reported.

Barbara McClintock discovered the first transposon, *Activator* / *Dissociation (Ac* / *Ds)* system, in maize in the late 1940s (McClintock 1948). During her continued work in the decades following the identification of *Ac* / *Ds*, she observed that “genomic stress” like breakage and rearrangement of chromosomes induced TE reactivation in maize. A number of kernels had dots of red or purple pigment, suggesting a silent *Dt* (“dotted”) element had been activated after the occurrence of a chromosome or chromatid type of breakage-fusion-bridge cycle (McClintock 1984). Maize En/Spm was discovered by Peter Peterson in the 1950s after gamma irradiation of maize kernels (Peterson 1953). Subsequently, reactivation of various TEs by different “genomic stress” events has been

reported. *Tto1* and *Tos17* retrotransposons are responsive to tissue culture in tobacco and rice, respectively (Takeda et al. 1999, Cheng et al. 2006). Wounding and elicitors induced activation of the tobacco retrotransposon *Tto1* is regulated by MYB-related transcription factor NtMYB2 (Takeda et al. 1998, Sugimoto et al. 2000). Transcriptional activation of the copia-like retrotransposon *Tnt1* can be induced by microbial elicitors, such as extracts of the fungus *Trichoderma viride* and bacterium *Erwinia chrysanthemi*, mechanical injury, chemical treatments and temperature change (Grandbastien et al. 1997). In *Antirrhinum*, *Tam3* is activated by low temperature (Hashida et al. 2003).

DNA damage is one stressor that can cause reactivation of transposable elements. In mammalian cells, SINEs are transcriptionally activated after exposure to cisplatin, etoposide, or gamma radiation (Rudin and Thompson 2001). In yeast, transposition of the *Ty1* retrotransposon is strongly activated by gamma-irradiation (Sacerdot et al. 2005). UV light, mitomycin C, and phleomycin induced DNA damages in the chromosome stimulate *Tn7* transposition in *Escherichia coli* (Shi et al. 2008). Both *mudrA* and *B* transcripts are expressed at higher levels after an 8h-UV-B treatment and accompanied by increased histone H3 acetylation and decreased DNA and H3K9me2 methylation in maize (Questa et al. 2010).

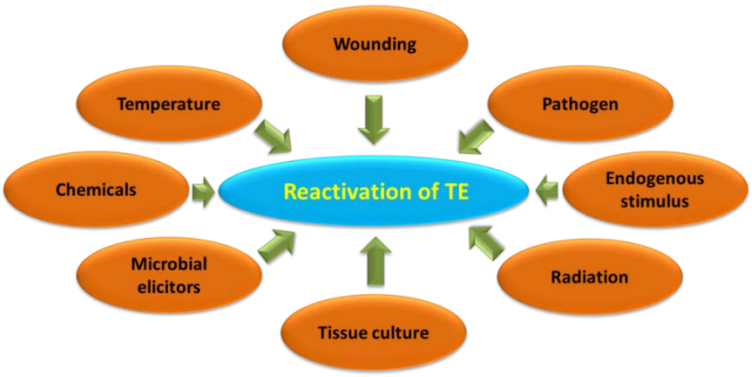


Figure 5: Silent TEs can be reactivated by “genomic stress”.

1.2 DNA damage repair is associated with epigenetic modifications

A number of reversion events of silent transposable elements induced by DNA damage were published (see previous chapter), but neither the regulatory pathway nor the

molecular machinery of this process are well understood. It is known, however, that activities of many transposable elements are regulated by epigenetic mechanisms. Recently, epigenetic regulation is found to play an important role in the DNA damage response as well, and it is possible that epigenetic regulation is the bridge between DNA damage response and activation of transposons.

1.2.1 Origins of DNA damage and repair pathways

DNA damage occurs during endogenous cellular processes and the exposure of organism to environmental stresses. The responses to damage are diverse and depend on the origin and type of lesion, such as single-strand damage and double-strand breaks (DSBs).

Reactive oxygen species (ROS) is a typical endogenous factor that induces the formation of oxidized bases, and further leads to single-strand breaks (SSBs) (Mitra et al. 2002, Hegde et al. 2008). UV light from sunshine mainly produces pyrimidine dimers on one strand of DNA (Tuteja et al. 2009). DNA damage at nucleotides activates base excision repair (BER), nucleotide excision repair (NER) or mismatch repair (MMR). BER is triggered by oxidation, alkylation, hydrolysis, or deamination induced single base damage. When UV-B light generates pyrimidine dimers, NER will take place in organism. MMR functions during the process of repairing errors of DNA replication and recombination (Charbonnel et al. 2011).

Replication errors occur spontaneously in cells and facilitate chromatid breaks (Lieber 2010). Environmental ionizing radiation, such as gamma irradiation and X-ray, bleomycin, and methylmethane sulfonate (MMS), can induce DSBs as well (Mannuss et al. 2012). There are three types of DSBs repair (DSBR) pathways: the non-homologous end joining (NHEJ) pathway, the homologous recombination (HR) repair pathway, and the microhomology-mediated end joining (MMEJ) pathway (Charbonnel et al. 2011).

1.2.2 Epigenetics in single-strand break repair

In recent years, histone modifications associated with single-strand break repair have been reported (Table 3). Short-lived transcription factor, p53, is able to mediate innate tumor suppression. The induction of p53 in response to DNA damage is coordinated by ATM and ATR induced phosphorylation. The phosphorylation of p53 stimulates the recruitment of key transcriptional proteins, such as p300 and CREB-binding protein (CBP), leading to the acetylation of several key lysine residues (Meek 2009). It has been reported that p53 and p300 maintain H3 acetylation in response to UV light in mammalian cells and *Drosophila* (Rebollar et al. 2006, Kim et al. 2009). In *Arabidopsis*, no homolog of p53 has been identified, but a plant specific protein, AtSOG1, plays a similar role (Yoshiyama et al. 2009). In maize, it has been observed that UV-B-tolerant lines exhibit greater acetylation on N-terminal tails of histones H3 and H4 after irradiation. These acetylated histones are enriched in the promoter and transcribed regions of *MBD101* and *NFC102* (Casati et al. 2008). After 4 hours of UV-B treatment, HAM1 and HAM2, members of *Arabidopsis* histone acetyltransferases of the MYST family, are induced in wild-type plants. However, cyclobutane pyrimidine dimers (CPDs) accumulation was higher in *ham1* than in *ham2*, suggesting that HAM1 has a major role in DNA repair after UV-B (Campi et al. 2012).

In *Saccharomyces cerevisiae*, Dot1p methylated H3K79 is required for nucleotide excision repair (NER) in response to UV (Bostelman et al. 2007, Chaudhuri et al. 2009). Inhibitor of growth 1 (ING1) defective cells are unable to activate UV induced damage repair because they lack the capacity to recognize H3K4me3 by its PHD finger (Pena et al. 2008).

Active DNA demethylation is executed by proteins of the DEMETER-LIKE DNA glycosylase family, which includes ROS1, DME, DML2 and DML3. After these 5-meC DNA glycosylases remove the methylated cytosine base and generate an abasic site, the gap is refilled with an unmethylated cytosine via base excision repair (BER) (Gong and Zhu 2011). Other DNA glycosylases, such as methylated DNA-binding protein (MBD4) or thymine DNA glycosylase (TDG) are associated with G/T mismatch repair

(MMR) in mammalian cells (Zhu 2009).

1.2.3 Epigenetics in double-strand breaks repair

In addition to single-strand break repair, the epigenetic machinery is also involved in DSBs (Table 3). Ionizing radiation (IR), such as gamma irradiation and X-ray, can cause DSBs in organisms. This damage is repaired by either HR or NHEJ (Jeggo 1998, Rothkamm et al. 2003). Before the repair process initiates either of these two above mentioned pathways, first the MRN complex (MRE11/RAD50/NBS1) recognizes the broken site, recruits Ataxia telangiectasia mutated (ATM) kinase and subsequently activates downstream repair signaling (Lee and Paull 2005). Tumour suppressor - Tip60, which is a histone and also ATM acetyltransferase, facilitates the formation of ATM-MRN complex via binding to H3K9me3 near a damage site (Sun et al. 2009). In *Arabidopsis*, *atm* mutants display increased histone acetylation upon X-ray irradiation (Drury et al. 2012). In *Drosophila*, a loss-of-function mutation of hMOF, a histone acetyltransferase that interacts with ATM, or RNAi-mediated hMOF knockdown blocks IR-induced increases in histone H4 acetylation at lysine 16 and leads to cell death, which results from decreased phosphorylation of downstream factors and DNA repair (Pandita and Richardson 2009).

After recognition of damage by MRN and recruitment of ATM, DNA end resection can occur which leads to the initiation of the HR pathway. Alternatively, the damaged site may be repaired via the NHEJ pathway. The selection is also dependent on the cell cycle phase, developmental stage and the cell type (Boyko et al. 2006, Shrivastav et al. 2008, Zierhut and Diffley 2008, Symington and Gautier 2011).

The 53BP1 protein prevents that DNA ends at breakage sites will be resected and, thus promotes the initiation of NHEJ repair (Bunting et al. 2010). 53BP1 plays a role in DNA repair in XRCC4-dependent pathways of DSB repair with the requirement of interaction with H4K20me2 in mammalian cells (Xie et al. 2007). Methylated H3K79 is also required for 53BP1 targeting to DSBs (Huyen et al. 2004). Ku70-Ku80 heterodimer recognizes the break site and subsequently recruits NHEJ components (Gospodinov

and Herceg 2013). The recruitment of Ku proteins is facilitated by CBP and p300 introduced H3 and H4 acetylation (Ogiwara et al. 2011). Besides histone acetylation and methylation, DNA methylation affects NHEJ repair as well. Ku80 expression is increased by 5-azacytidine-induced demethylation and enhances retinal neurocytes DNA repair (Zhuang et al. 2010).

Table 3: DNA damage related epigenetic events

Type	Species	Conditions	Reference
BER	*A. thaliana*	DNA demethylation	(Gong and Zhu 2011)
BER	*Z. mays*	H3,H4 acetylation	(Casati et al. 2008)
BER	*A. thaliana*	Histone acetylation	(Campi et al. 2012)
NER & NHEJ	*H. sapiens*	H3,H4 acetylation	(Meek 2009, Ogiwara et al. 2011)
NER&DSBR	*S. cerevisiae, H. sapiens*	H3K79 methylation	(Bostelman et al. 2007, Chaudhuri et al. 2009, Sun et al. 2009)
NER	*H. sapiens*	H3K4me3	(Pena et al. 2008)
DSBR	*A. thaliana, H. sapiens*	Histone acetylation	(Sun et al. 2009, Drury et al. 2012)
DSBR	*D. melanogaster*	H4K6 acetylation	(Pandita and Richardson 2009)
DSBR	*H. sapiens*	H3K4me3	(Li et al. 2014)
NHEJ	*H. sapiens*	DNA methylation	(Zhuang et al. 2010)
NHEJ	*H. sapiens*	H3K36me3	(Fnu et al. 2011)
NHEJ	*H. sapiens*	H4K20me2	(Xie et al. 2007)
HR	*H. sapiens*	DNA demethylation	(Cuozzo et al. 2007)
HR	*H. sapiens*	H3K36me2, H3K36me3,H3K4me1	(Daugaard et al. 2012)
HR	*H. sapiens*	H3K9ac	(Aymard et al. 2014)

BER: base excision repair; NER: nucleotide excision repair; DSBR: DSBs repair, includes non-homologous end joining (NHEJ) and homologous recombination (HR) repairs.

The HR pathway requires breast cancer susceptibility 1 (BRCA1) promoted resection of DNA ends at damage sites (Schlegel et al. 2006, Yun and Hiom 2009). Lysine-specific histone demethylase 5B (KDM5B) demethylates H3K4me3 at DNA damage sites, which in turn, facilitates both NHEJ and HR repair by recruiting Ku70 and BRCA1 to the breaks (Li et al. 2014). BRCA1 initiates HR repair with the help of C-terminal binding protein interacting protein (CtIP) (Yun and Hiom 2009). While CtIP binds Lens

epithelium–derived growth factor p75 splice variant (LEDGF), which interacts with H3K36me2, H3K36me3, and H3K4me1, in turn, this subsequently triggers RAD51 functions (Daugaard et al. 2012). RAD51 binds to damage sites and accompanies enriched H3K36me3 and H3K9ac (Aymard et al. 2014).

1.3 Chromatin remodeling

Epigenetic modifications including DNA methylation, histone methylation and acetylation are associated with DNA damage response and regulation of TEs. Chromatin remodeling, one of several other processes involved in DNA damage repair (DDR), is necessary to allow repair factors gain access to lesions. In yeast and mammalian cells, the SWI2/SNF2 ATPase domain superfamily chromating remodeling proteins have been shown to function in response to NER and DSB repair (Lans et al. 2012). The *Arabidopsis SWI2/SNF2* gene family has 40 members. Their involvement in DNA recombination and repair has been demonstrated, e.g. 14 of them respond to gamma or UV radiation induced DNA damage (Shaked et al. 2006). SW1/SNF stimulation of BER is reported as well (Menoni et al. 2007). Chromatin remodeler, INO80 protein, promotes HR via resecting break ends (Gospodinov et al. 2011). Four maize chromatin genes, *CHC101*, *NFC102*, *SDG102*, and *MBD101*, are illustrated to be important during DNA repair by UV-B irradiation (Campi et al. 2012).

So far, little is known about chromatin remodeling regulated TEs. However, commonly used factors, such as SWI2/SNF2 proteins, suggest chromatin remodeling functions potentially in both DNA damage response and regulating TEs. By the induction of UV-B, *mutator* elements are activated and *MuDR* TIRs are enriched in SWI2/SNF2 ChIP assay, which indicated the involvement of chromatin remodeling (Questa et al. 2010). Depletion of nucleosome remodeling factor complex component101 (NFC101) and NFC102 transcriptionally activates TEs in maize (Mascheretti et al. 2013). In plants, transposons are silenced mainly by non-CG methylation and H3K9 methylation, which requires the chromatin-remodeling gene *DDM1* (Sasaki et al. 2012). 5-aza-2‘-deoxycytidine (5-Aza-CdR) and 4-phenylbutyric acid (PBA) induced chromatin

remodeling at *Alu* elements results in microRNA-512-5p upregulation in human cells (Saito et al. 2009).

1.4 Mis-processing of *Ac* transcripts in *Arabidopsis*

Maize *Ac*, the first identified transposable element by McClintock in the late 1940s, belongs to the *hAT* transposon family. It encodes an 807 amino acids transposase and moves via DNA intermediates by a “cut-and-paste” mechanism. Since the first demonstration that *Ac* is capable of autonomously transposing in a heterologous plant, tobacco (Baker et al. 1986), *Ac* and non-autonomous *Ds* elements have been introduced and shown to be functional in more than 20 other plant species (Kunze and Weil 2002), including tomato (Yoder et al. 1988), rice (Izawa et al. 1991), Lotus (Thykjaer et al. 1995) and *Arabidopsis* (Van Sluys et al. 1987). It was also reported that *Ac/Ds* are able to transpose in yeast (Weil and Kunze 2000) and vertebrate cells (Emelyanov A 2006) which indicates the versatile functions of *Ac/Ds* in eukaryotes.

In maize and several heterologous plant species the *Ac/Ds* system has been proven to be a powerful genetic tool for insertional mutagenesis and gene tagging, such as in *Arabidopsis*, rice, maize, and barley (Altmann et al. 1992, Athma et al. 1992, Grevelding et al. 1992, Koprek et al. 2000, Greco et al. 2001, Kim CM 2004, Froschauer et al. 2012). Recently, *Ac/Ds* elements were used for discovering genes in poplar (Fladung and Polak 2012).

Although *Ac* is able to transpose in many, if not all plants, the observed transposition frequencies vary significantly in different species. *Ac* transposition frequencies in *Arabidopsis,* as well as in many important agricultural crops, are typically 1 % or less, which is lower than in maize and tobacco. Thus, it is too low to be efficiently used as a genetic tool. The low transposition frequencies are a result of the complex regulation of *Ac* by DNA methylation, negative dosage effect, and possible inhibitor function of truncated TPase proteins (McClintock 1949, McClintock 1951, Schwartz and Dennis 1986, Scofield et al. 1993, Boehm et al. 1995, Wang et al. 1996).

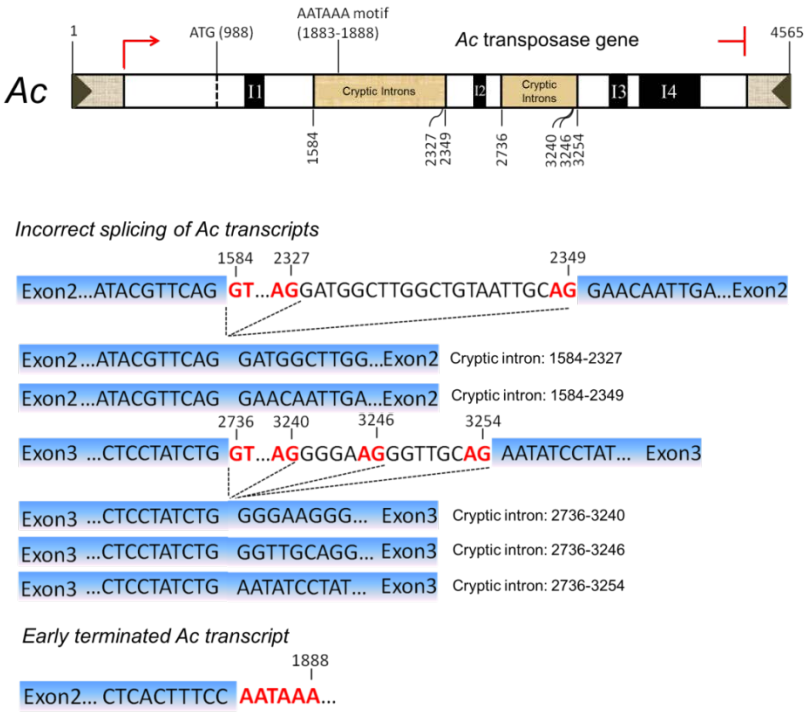


Figure 6: Positions of *Ac* mis-processed transcripts detected in *Arabidopsis* and sugar beet.

Top diagram: Structure of *Ac*. Full Ac DNA sequence is 4565 bp long. I1-I4: introns 1-4. Red right arrow: transcription start site. ATG: AcTPase start codon at 988. AATAAA motif at position 1883-1888: leads to early terminated transcripts. Red terminating marker: Polyadenylation site of correct *Ac* transcript. Dark triangles: terminal inverted repeats. Brown frames: cryptic introns. Numbers flanking brown frames: starting and ending positions of cryptic introns, details see down diagram. Cryptic intron 2736-3240 was detected in both *Arabidopsis* and sugar beet. The other cryptic introns were only detected in sugar beet. Deletions starting at position 1584 can stop at either 2327 or 2349. Deletions starting at position 2736 end at 3240, 3246 or 3254. All cryptic introns follow the GT-AG rule of intron splicing. Red AATAAA motif: start region of early terminated *Ac* transcripts detected in *Arabidopsis*.

One possible reason for low transposition frequencies might be aberrant splicing of the *Ac* primary transcript in the heterologous host, which was reported for *Arabidopsis* and sugar beet. In transgenic *Arabidopsis* plants seven different transcripts were found, six of which resulted from premature termination and polyadenylation at sites between nucleotides 1883 and 1918 within exon 2 with AATAAA motif (1883 -1888) (Figure 6) (Jarvis et al. 1997). The seventh clone analyzed showed correct splicing of introns 1, 2, 3 and 4 but carried a 505 bp deletion within exon 3, which they interpreted to represent a cryptic intron (Figure 6) (Jarvis et al. 1997). Recently, Lisson and colleagues introduced (*Ac/Ds*) transposable element into sugar beet. They found that the fourth

intron of the transposase gene was partially mis-spliced. Four different splice products were identified. In addition, the second and third exons were found to harbor two and three novel introns, respectively (Figure 6) (Lisson et al. 2010).

One possible explanation for mis-processing of heterologous TE transcripts could be that during phylogenesis the plants were continuously or repeatedly exposed to foreign transposons, and thus have evolved such a defense mechanism. However, so far, there is no data that support this hypothesis. It is more likely that in different plants the splicing machinery has somewhat different affinities to intron/exon border sequence compositions, so if this composition differs in maize, aberrant splicing may be the consequence.

1.5 RNA-seq technology

In plants, several studies addressed gene regulation and transcriptome changes in response to DSBs (Doucet-Chabeaud et al. 2001, Gallego et al. 2001, Culligan et al. 2006, Ricaud et al. 2007, Chen et al. 2008), but the pathways are still not well understood. In order to understand the genomic response to DDR mechanisms, several methods, for example, quantitative RT-PCR (qRT-PCR) and microarrays have been used in previous studies (Culligan et al. 2006, Ricaud et al. 2007).

Recently, a cDNA based high-throughput sequencing technology, called RNA-seq, was developed for transcriptomic analysis. A library of cDNA fragments with adaptors attached to one or both ends is generated from mRNA. Single end reads (from one end) or paired-end reads (from both ends) are obtained by high-throughput sequencing from the molecules with adaptors. Depending on the sequencing platform used for RNA-seq, e.g. Illumina HiSeq, Applied Biosystems SOLiD or Roche 454/GS, the read lengths can vary from 30 to 1000 bp (Figure 7) (Wang et al. 2009). According to the purposes and the materials used for sequencing, different alignment tools, for example, Tophat (Trapnell et al. 2012), SOAP (Li et al. 2008, Li et al. 2009b), RobiNA (Lohse et al. 2012), or bowtie (Langmead et al. 2009, Langmead and Salzberg 2012), can be selected to

carry out *de novo* assembly without knowledge of the genomic sequence, to align reads to a reference genome, or to map reads to reference transcripts. Although RNA-seq meanwhile has been used widely for transcriptome analysis in different species, it is still a fast developing technique.

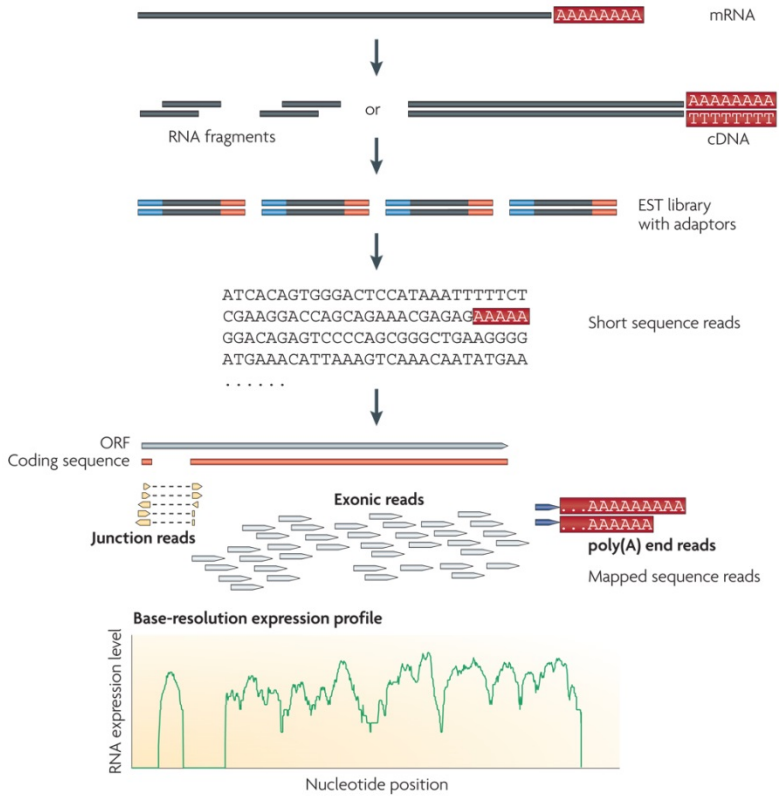


Figure 7: A typical RNA-seq experiment (Wang et al. 2009).

The mRNAs are poly(A) enriched from total RNA, then a cDNA library with adaptors at either single-end or paired-ends is generated. Reads are produced by different sequencing platforms, such as Illumina HiSeq, Applied Biosystems SOLiD or Roche 454/GS. The generated junction reads, exonic reads and poly(A) end-reads are aligned to reference genome to create expression information of each gene, discover novel transcripts and/or alternative splicing of genes (Wang et al. 2009). Reads generated from species without known genomic information can also be used for creating expression profile of genes with *de novo* assembly and further analysis.

TEs are not well represented on the commercial *Arabidopsis* microarrays. Therefore, for purpose of this work, RNA-seq is considered to be the superior technique for detecting regulated TEs after DNA damage. So far, some groups have identified TEs/TE-related sequences successfully by RNA-seq. Upregulated transposons and intergenic transcripts in the *Arabidopsis* DNA demethylation-deficient rdd mutant were detected by

RNA-seq and combined with methylC-seq to generate an integrated epigenome map of *Arabidopsis* (Lister et al. 2008). RNA-seq contributed to the detection of 2708 lincRNAs (long intergenic non-coding RNAs) expressed in *Arabidopsis*, and among these lincRNAs, TEs/TEs-related transcripts were observed (Liu et al. 2012). By using RNA-seq, it was observed that 78 % of differentially expressed DNA transposons were upregulated, while, 68 % of differentially expressed retrotransposons were downregulated in *RDR2* defective plants (Jia et al. 2009). SINE repeat-derived Pol V-dependent small RNAs were recently analyzed by RNA-seq (Lee et al. 2012). RNA-seq revealed expression of 105 retroelements in locust (Jiang et al. 2012). These examples illustrate the usefulness of RNA-seq for the purpose of this study.

1.6 Goals of the research

How TEs are negatively regulated and eventually silenced by the epigenetic machinery is known in some detail. The phenomenon that epigenetically silenced TEs can be reactivated in response to DNA damage has also been known for years. However, the process of reactivation of silent TEs is still poorly understood. In order to shed light on this field, double-strand breaks (DSBs) will be generated in *A. thaliana* plants to investigate the consequences on the transcriptome with the emphasis on transposons.

Although activation of silenced TEs was observed in previous studies, these studies focused only on individual TEs. A major aim of this work is a more comprehensive investigation of TE reactivation, particularly by the induction of DSBs. It will help to answer the following questions: Are TEs collectively up- or downregulated or just occasionally and randomly after DNA damage? Another open question is whether DNA transposons and retrotransposons are similarly regulated. In addition to responses of TEs by the induction of DSBs, it is worth investigating if DSBs repair (DSBR) genes regulate reactions of TEs. DNA damage sensor genes or highly regulated DDR genes will be good candidates for answering this question.

Epigenetic machinery controls both TEs and DSBR. Another task of this work is to answer the question whether TEs in response to DSBs are epigenetically regulated. First, epigenetic genes induced by DSBs will be identified. Subsequently, mutant lines that lack those epigenetic genes will be used for investigating epigenetic regulation of TEs in response to DSBs. Furthermore, analyzing alterations of DNA methylation at target loci and histone modifications after DNA damage also can improve the understanding of epigenetic machinery in this process.

In parallel to the above mentioned work, to prevent mis-processed *Ac* transcripts, previously reported cryptic introns and early terminated sites will be point mutated for creating modified *AcTPase* coding sequences (CDSs) without changing amino acids of *Ac* transposase. This part of the work is an additional replenishment for investigating the regulation of heterologous TEs after being introduced into *Arabidopsis*.

2 Materials and Methods

2.1 Materials

2.1.1 Antibodies, Enzymes and Kits

Antibodies used in this work are listed in table 4.

Table 4: Antibodies applied

Antibodies	**Companies**
Goat Anti-Mouse IgG-HRP	Santa Cruz Biotechnology (Santa Cruz, CA, USA)
Goat Anti-Rabbit IgG-HRP	Santa Cruz Biotechnology (Santa Cruz, CA, USA)
Anti-Histone H3 antibody	Abcam plc (Cambridge, UK)
Anti-Dimethyl Histone H3 (Lys4)	EMD Millipore (Billerica, MA, USA)
Anti-Histone H3 (tri methyl K4)	Abcam plc (Cambridge, UK)
Anti-Histone H3 (acetyl K9)	Abcam plc (Cambridge, UK)
Anti-Histone H3 (di methyl K36)	Abcam plc (Cambridge, UK)

All restriction enzymes were purchased from either Thermo Fisher Scientific (Bonn, Germany) or New England Biolabs (Frankfurt, Germany).

Without special indication, self-made *Taq* DNA polymerase was applied to all regular PCR reactions. Preparation of self-made *Taq* DNA polymerase is according to Desai's work (Desai and Pfaffle 1995). Commercial *Taq* DNA polymerase (Thermo Fisher Scientific, Schwerte, Germany) was used in case a higher efficiency was required. Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Schwerte, Germany) was applied for the accurate amplification of DNA.

Other commonly used enzymes were purchased from the companies listed as bellow:

Life Technologies (Darmstadt, Germany)

New England Biolabs (Frankfurt, Germany)

Thermo Fisher Scientific (Schwerte, Germany)

All commercial kits used were listed in table 5.

Table 5: Kits applied

Kits	Companies
CloneJET^{TM} PCR Cloning Kit	Thermo Fisher Scientific (Schwerte, Germany)
DNeasy Plant Mini Kit	QIAGEN (Hilden, Germany)
EpiTect Fast Bisulfite Conversion Kits	QIAGEN (Hilden, Germany)
NucleoSpin® Gel and PCR Clean-up	MACHEREY-NAGEL (Düren, Germany)
Pierce ECL Western Blotting Substrate	Thermo Fisher Scientific (Schwerte, Germany)
QuikChange II Site-Directed Mutagenesis Kit	Agilent Technologies, Santa Clara, CA, USA
RNeasy Mini Kit	QIAGEN (Hilden, Germany)

2.1.2 Ladders

λ-PstI-DNA-Marker and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Schwerte, Germany) were used for measuring the sizes of DNA fragment on agarose gel.

The sizes of proteins were checked by comparing with Prestained Protein Marker, Broad Range (7-175 kDa) (New England Biolabs, Frankfurt, Germany).

2.1.3 Oligonucleotides

All oligonucleotides were synthesized by Life Technologies (Darmstadt, Germany). Oligonucleotides are listed in tables according to their different usages (Table 6), detail information check tables in appendix.

Table 6: Tables of primers listed in appendix

Table names	Contents
Table S1	Primers for generating constructs of estradiol induced cleavage system
Table S2	Primers for *Ac*
Table S3	Primers for qRT-PCR
Table S4	Primers for T-DNA insertion lines
Table S5	Primers for Bisulfite sequencing analysis

2.1.4 Plasmids

Plasmids used and generated in this work are in table 7.

Table 7: Plasmids list

Plasmid names	**Descriptions and references**	**Selection markers in Bacteria (B), in Plants (P)**
p35S-I-Sce I	I-Sce I CDS (Puchta 1999b)	AMP^R (B)
pCAMBIA1302	Binary T-DNA vector with cauliflower mosaic virus (CaMV) 35S promoter (CambiaLabs)	KAN^R (B) HYG^R (P)
PET28.XH.QQR	QQR CDS (Tovkach et al. 2009)	KAN^R (B)
pET28.XH.ZFN3	ZFN3 CDS (Tovkach et al. 2009)	KAN^R (B)
pMDC7	Binary T-DNA vector for expression from an estrogen inducible promoter (Curtis and Grossniklaus 2003)	$Spec^R$ (B) HYG^R (P)
pMK-RQ-Ac	Synthesized *Ac* mutation fragment (Mr.Gene)	KAN^R (B)
pKU2	2'-p-1'::KAN fragment (Baker et al. 1987)	AMP^R (B)
pRK19	AcTPase(1-807) (Prof. Reinhard Kunze)	AMP^R (B)
pSAT6A.QQR-TS*::GUS	GUS CDS interrupted by QQR recognition site (Tovkach et al. 2009)	AMP^R (B)
pSAT6A.ZFN3-TS*::GUS	GUS CDS interrupted by ZFN3 recognition site (Tovkach et al. 2009)	AMP^R (B)
pVC-SCB695-1qcz	*Arabidopsis* codon-optimized I-Sce I gene with an intron under control of the parsley ubiquitin promoter (SunGene)	KAN^R (B)
pZW-O1	Binary T-DNA vector contains GUS CDS interrupted by I-Sce I recognition site	$Spec^R$ (B) KAN^R (P)
pZW-O2	Binary T-DNA vector contains GUS CDS interrupted by ZFN3 recognition site	$Spec^R$ (B) KAN^R (P)
pZW-O3	Binary T-DNA vector contains GUS CDS interrupted by QQR recognition site	$Spec^R$ (B) KAN^R (P)
pZW-O4	Binary T-DNA vector contains *I-Sce I* CDS downstream of an estrogen inducible promoter	$Spec^R$ (B) HYG^R (P)
pZW-O5	Binary T-DNA vector contains ISI-opA (*Arabidopsis* codon-optimized *I-Sce I* gene) CDS downstream of an estrogen inducible promoter	$Spec^R$ (B) HYG^R (P)
pZW-O6	Binary T-DNA vector contains *ZFN3* CDS	$Spec^R$ (B)

	downstream of an estrogen inducible promoter	HYG^R (P)
pZW-O7	Binary T-DNA vector contains *QQR* CDS	$Spec^R$ (B)
	downstream of an estrogen inducible promoter	HYG^R (P)
pZW-O8	Binary T-DNA vector contains *I-Sce I* CDS	$Spec^R$ (B)
	downstream of a 35S promoter	HYG^R (P)
pZW-O9	Binary T-DNA vector contains *ISI-opA* CDS	SpecR (B)
	downstream of a 35S promoter	HYGR (P)
pZW-O10	Binary T-DNA vector contains *ZFN3* CDS	SpecR (B)
	downstream of a 35S promoter	HYGR (P)
pZW-O11	Binary T-DNA vector contains *QQR* CDS	SpecR (B)
	downstream of a 35S promoter	HYGR (P)
pZW-O12	Binary T-DNA vector contains *AcTPase* (1-807)	SpecR (B)
	CDS with 3 point mutations (a1886c, t2737a, and a3239c) downstream of a 35S promoter	HYGR (P)
pZW-O13	Binary T-DNA vector contains *AcTPase* (1-807)	SpecR (B)
	CDS with 7 point mutations (a1886c, g2327a, a2348c, t2737a, a3239c, a3245c, and a3253c) downstream of a 35S promoter	HYGR (P)
pZW-O14	MiniDs was inserted between HindIII in pZW-O1 by SLIC	$Spec^R$ (B) KAN^R (P)
pZW-O15	MiniDs was inserted between HindIII in pZW-O2 by SLIC.	$Spec^R$ (B) KAN^R (P)
pZW-O16	MiniDs was inserted between HindIII in pZW-O3 by SLIC.	$Spec^R$ (B) KAN^R (P)
pZW-O20	Binary T-DNA vector contains WT *AcTPase* (1-807) CDS downstream of a 35S promoter	SpecR (B) HYGR (P)

2.1.5 Bacteria strains

Escherichia coli and *Agrobacterium tumefaciens* strains used in this work are:

Escherichia coli

XL1-Blue Stratagene (La Jolla, CA, USA)

recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1

lac [F' *proAB lac*IqZ" M15::Tn10 (Tetr)]

DH 10-beta

F- mcrA ” (mrr-hsdRMS-mcrBC) ¦ 80lacZ” M15 endA1 recA1 nupG rpsL ” lacX74 araD139 ” (ara,leu)7697 »-

Agrobacterium tumefaciens

GV3101::pMP90 (Koncz and Schell 1986)

Rifr, Gentr

2.1.6 Plant material

Arabidopsis thaliana WT (Col-0) and mutant plants: *atm-2* (SALK_006953), *ATBRCA1-1* (SALK_014731), *drm1-2* (SALK_021316), *drm2-2* (SALK_150863), *jmjd5-1* (SAIL_811_H12), and *ago4-5* (WISCDSLOX338A06) were obtained from NASC. Homozygous plants *ago2-1* (SALK_003380) and *ago7-2* (SALK_095997) were gifts from Prof. Sir David Baulcombe.

The miniDs 3-44-1 line with a single inserted T-DNA fragment, which contains a miniDs element inserted between a 35S promoter and an eGFP-Bar fusion gene, was obtained from My-Linh Doll (Doll 2008).

Transgenic plants generated by plant transformation are listed in Table 8.

Table 8: List of generated transgenic plants

***Arabidopsis* lines**	**Descriptions**
miniDs-ISceI-TS::GUS	Col-0 transformed with pZW-O14 construct
miniDs-ZFN3-TS::GUS	Col-0 transformed with pZW-O15 construct
miniDs-QQR-TS::GUS	Col-0 transformed with pZW-O16 construct
XVE-ISceI	Col-0 transformed with pZW-O4 construct
XVE-ISI-opA	Col-0 transformed with pZW-O5 construct
XVE-ZFN3	Col-0 transformed with pZW-O6 construct
XVE-QQR	Col-0 transformed with pZW-O7 construct
35S::ISceI	Col-0 transformed with pZW-O8 construct
35S::ISI-opA	Col-0 transformed with pZW-O9 construct
35S::ZFN3	Col-0 transformed with pZW-O10 construct
35S::QQR	Col-0 transformed with pZW-O11 construct
35S::cAc-M3	Col-0 transformed with pZW-O12 construct
35S::cAc-M7	Col-0 transformed with pZW-O13 construct
35S::cAc	Col-0 transformed with pZW-O20 construct

2.1.7 Gene synthesis

Artificial *AcTPase* CDS (1173-2421, ORF) containing 6 point mutations (g2327a, a2348c, t2737a, a3239c, a3245c, and a3253c) was synthesized by MR.GENE (Regensburg, Germany).

2.1.8 Sequencing

DNA sequencing was done by GATC Biotech (Konstanz, Germany) and LGC Genomics (Berlin, Germany).

RNA sequencing was done by GATC Biotech (Konstanz, Germany) and BGI-Tech (Hong Kong, China).

2.1.9 Computer programs and database

Computer programs used were collected in table 9 and 10.

Table 9: Commonly used computer programs

Computer programs	**References**
CorelDRAW Graphics Suite X6	Corel (Ottawa, Canada)
FileMaker Pro 8	FileMaker (Unterschleißheim, Germany)
OriginPro 8.6	OriginLab (Northampton, MA, USA)
SnapGene Viewer 2.2.1	GSL Biotech LLC (Chicago, IL, USA)
Vector NTI Advance® 11.5	Life Technologies (Darmstadt, Germany)
SPSS21	IBM (Armonk, New York, USA)

Table 10: List of programs for RNA-seq analysis

Names	**Sources**	**References**
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	(Langmead and Salzberg 2012)
Tophat	http://tophat.cbcb.umd.edu/	(Trapnell et al. 2012)
Cufflinks	http://cufflinks.cbcb.umd.edu/	(Trapnell et al. 2012)
CummeRbund	http://compbio.mit.edu/cummeRbund/	(Trapnell et al. 2012)
Mapman	http://mapman.gabipd.org/	(Thimm et al. 2004)
IGB	http://arabidopsis.org/doc/tools/igb/91	(Nicol et al. 2009)
R studio	http://www.rstudio.com/	

Table 11: Online tools

Computer programs	Sources	References
CyMATE	http://www.cymate.org/	(Hetzl et al. 2007)
agriGO	http://bioinfo.cau.edu.cn/agriGO/	(Du et al. 2010)
Kismeth	http://katahdin.mssm.edu/kismeth/primer_design.pl	(Gruntman et al. 2008)
QuantPrime	http://www.quantprime.de/main.php?page=home	(Arvidsson et al. 2008)

Sequences and literatures were obtained from National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/; Be-thesda, MD, USA). *Arabidopsis* genome sequences and annotated gene model (version: TAIR10) were downloaded from tophat supplied sources (http://tophat.cbcb.umd.edu/igenomes.shtml). Transposable element annotated models and sequences were obtained from the *Arabidopsis* Information Resource (TAIR) (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_transposable_elements/). Sequence information of long non-coding RNAs was searched via PLncDB (Plant Long noncoding RNA Database, http://chualab.rockefeller.edu/gbrowse2/homepage.html).

2.2 Methods

2.2.1 Handling bacteria and plants

2.2.1.1 *Escherichia coli* transformation and culture

50 µl *Escherichia coli* (*E. coli*) competent cells were thawed on ice, and then mixed with 5 µl ligated product or required amount of target DNA mixture. After 30 min incubation on ice, cells were heat shocked at 42 °C for 45 sec, chilled on ice for 2 min, and then mixed with 450 ml pre-warmed SOC (2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose, PH 7.5) or LB (1 % tryptone, 0.5 % yeast extract, 1 % NaCl, PH 7.0) and incubated at 37 °C for 1 hour while shaking at 220 rpm. Plated 200 µl medium on LB plates within demanded antibiotics,

incubated at 37 °C overnight.

Inoculating *E. coli* in liquid LB was done under the same conditions as culturing bacteria on LB plates, while shaking at 220 rpm.

2.2.1.2 *Agrobacterium tumefaciens* transformation and culture

After having been thawed on ice, 1 µl plasmid DNA was added to 50 µl *A. tumefaciens* cells, mixed briefly and incubated on ice for 10 min. The mixture was then transferred to a pre-chilled cuvette, and then the surround of the cuvette was wiped to remove water. The cuvette was put in an electroporator, and shocked with 2.2 kV for 5 ms. The electroporated A. tumefaciens cells were transferred into a new tube containing 1.9 ml SOC or LB medium and then incubated at 28°C for 2 hours at a speed of 220 rpm. 20 µl – 50 µl medium were plated on LB plates within required antibiotics and incubated at 30°C for 2 days.

A. tumefaciens in liquid LB was cultured at 28 °C while shaking at 220 rpm for 1 - 2 days depending on the usage.

2.2.1.3 *A. thaliana* growth conditions

A. thaliana seeds were sterilized before grown on half MS medium plates (Murashige and Skoog 1962). Seeds were treated with 70 % ethanol for 2 min, 10 % NaClO and 1 % SDS for 3 min, washed 3 times in autoclaved double deionized water for 3 min, and plated on plates with 0.1 % Agarose or dried with filter paper then put on plates with the required numbers of seeds. Plates sealed with Parafilm stayed at 4 - 8 °C for 2 days and were then transferred to an artificial climate chamber or light room under either long day (16 hours light/d) or short day (8 hours light/d) conditions (Weigel and Glazebrook 2002).

Seeds sown in soil in pot (6 cm ø or 12 cm ø) were first incubated at 4 - 8°C for 2 days. 10 days after germination seedlings were singled out into 6 cm ø pots for further cultivation in the greenhouse under long day conditions. Seeds harvesting and further

growth conditions were described in the manual (Weigel and Glazebrook 2002)

2.2.1.4 *Arabidopsis thaliana* transformation

Arabidopsis plants were transformed with constructs described above by the floral-dip-method (Clough and Bent 1998).

Overnight inoculated 200 ml *A. tumefaciens* culture were spun down at 4000 rpm at room temperature. The pellet was resuspended in freshly prepared infiltration medium (half MS with 44 nM BAP, 0.005 % Silwet L-77 [v/v], and 5 % sucrose [w/v], pH 5.8). 4 single plants or a big pot with approximately 20 plants were dipped in *A. tumefaciens* medium. Infiltrated plants were then moved back to the greenhouse and grown until seed harvesting.

2.2.1.5 Crossing plants

The crossing operation was carried out under a ZEISS microscope. Immature anthers were released from buds just before flowering. Mature stamen from male crossing plants were pollinated with pistils from female crossing plants. Crossed plants were then grown in the greenhouse under long day conditions until seeds were harvested.

2.2.1.6 GUS staining

Seedlings were put in 2 ml GUS-staining solution (0.1 % X-Gluc, 1 % Triton X-100, and 50 mM phosphate buffer pH7.2) and vacuum infiltrated for 2 min in Speed-Vac. After incubation at 37 °C overnight or for a specified time, seedlings were washed 3 times at RT with ethanol and acetic acid mixture (3:1, v/v) (Jefferson et al. 1987).

2.2.2 Molecular biological methods

Methods of general molecular biological experiments followed the instructions in Molecular cloning (3rd edition) (Sambrook and Russell 2001).

2.2.2.1 Site-directed mutagenesis of *AcTPase* CDS

QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) was used for generating point mutations. Mutagenesis primers were designed according to the Primer Design Guidelines of the kit.

Components in a 50 µl mutagenesis PCR reaction were: 1x reaction buffer, 50 ng plasmid template (pRK19), 125 ng each mutagenesis primer, 1 dNTP mix (10 mM total), and 2.5 U *PfuUltra* HF DNA polymerase. Amplification was carried out as: 95 °C 30 s, (95 °C 30 sec, 55 °C 1 min, 68 °C 6 min) x 12 cycles. After 2 min chilled on ice, 1 µl *DpnI* was used for removing the template DNA by incubation at 37 °C for 1 hour. 1 µl digested product was then added to 50 µl XL-1-Blue chemical competent cells for transformation (details of transformation see 2.2.1.1).

2.2.2.2 Constructs generation

All DNA fragments used for generating constructs were amplified by using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Schwerte, Germany). 2 tubes of 50 µl PCR were applied for amplifying each DNA fragment.

Nuclease recognition sites constructs:

Sequence and Ligation-independent Cloning (SLIC) (Li and Elledge 2007) was used to generate the constructs.

Fragments containing GUS CDS interrupted by recognition sites of ZFN3 and QQR were amplified from PSA6A.QQR-TS::GUS and PSA6A.ZFN3-TS::GUS, respectively. I-Sce I target site was added by replacing the ZFN3 target site by using PSA6A.ZFN3-TS::GUS as a template for PCR. 2'-p-1'::KAN fragment was obtained by PCR from pKU2. All PCR products were digested with *DpnI* (1 µl *Dpn* I / 100 µl product) at 37 °C for 1 hour after amplification. *pMDC7* was digested with *BamH* I and *Spe* I as the backbone. All *Dpn* I treated fragments and the digested backbone were harvested by column purification. 1µg DNA from each sample was treated with T4 DNA

polymerase (NEB) at room temperature for 30 min, to create overhangs.

pZW-O1 was generated by an annealing reaction which assembled 2'-p-1'::KAN, I-Sce I-TS::GUS, and backbone fragments (0.05277 pmol of each fragment) in the presence of *RecA* (1µg) and incubation at 37 °C for 30 min. pZW-O2 and pZW-O3 were created by the same strategy as described above, but by replacing their recognition fragments (ZFN3-TS::GUS and QQR-TS::GUS) respectively.

Nucleases constructs:

CDSs of *I-Sce* I, *ISI-opA, ZFN3*, and *QQR* were amplified from PET28.XH.QQR, p35S-I SceI, and pVC-SCB695-1qcz respectively. *Dpn* I digestion of PCR products was carried out directly after amplification. As a backbone, *pMDC7* was digested with *Asc* I and *Spe* I. After column purification, the DNA was treated with T4 DNA polymerase, and the annealing reaction was done at the presence of *RecA*. Constructs pZW-O4 *(I-Sce* I*),* pZW-O5 *(ISI-opA),* pZW-O6 *(ZFN3), and* pZW-O7 *(QQR)* were accomplished by inserting CDSs between *Asc* I and *Spe* I on pMDC7 via SLIC.

AcTPase constructs:

Three ORF segments were amplified from previously point mutated pRK19 constructs. Each contains one point mutation, a1886c, t2737a, and a3239c respectively, and was then assembled between *Nco* I and *BstE* II sites of pCambia1302 in an order of *Nco* I-a1186c-*Nar* I, *Nar* I-t2737a-*Sph* I, and *Sph* I-a3239c-*BstE* II, and finally generated pZW-O12.

pZW-O13 was generated via assembling two ORF fragments by SLIC. *Ac* ORF fragment 'a' including one mutation (a1886c) was obtained by PCR using point mutated pRK19 (a1886c) as template. *Ac* ORF fragment 'b' containing six point mutations (g2327a, a2348c, t2737a, a3239c, a3245c, and a3253c) was amplified from *de novo* synthesized *Ac* DNA sequence generated by MR.GENE. Those two segments were inserted between *Nco* I and *BstE* II sites in pCambia1302 as well.

pZW-O20 was created by inserting the full length wild type (WT) Ac transposase ORF, cleaved from pRK19 with *Nco* I and *Sma* I, in pCambia1302 between *Nco* I and *BstE* II

sites.

2.2.2.3 Genomic DNA isolation

Genomic DNA for normal usage and southern blot was extracted by CTAB protocol (Murray and Thompson 1980).

Arabidopsis seedlings or leaves were ground with liquid nitrogen by mortar or Retch® MM400. 600 µl Buffer B (100 mM Tris-Cl pH 8.0, 1.4 M NaCl, 20mM EDTA, and 2 % CTAB) pre-warmed at 65 °C was added to material powder and mixed well. The tubes were incubated at 65 °C for 20 min in a water bath, and inverted every 5 min. The clear supernatant was transferred to a new tube after centrifuge at room temperature (RT) for 5 min 13500 rpm. One volume chloroform / isoamyl alcohol (24:1) was added to the tube and mixed by inverting, then centrifuged at 13500 rpm for 5 min. The upper phase was then transferred into a fresh tube and repeat chloroform / isoamyl alcohol extraction was repeated once. One volume Buffer C (50 mM Tris-Cl pH 8.0, 10 mM EDTA, and 1 % CTAB) and one volume isopropanol were added to the supernatant, then mixed and incubated at room temperature for 10 min. It was then centrifuged at 12000 rpm for 5 min to precipitate DNA. The pellet was resuspended in 400 µl STE (10 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM EDTA, 20 µg/ml RNaseA) and incubated at 65 °C for 5 min. The DNA was pelletized in 600 µl ethanol at 13500 rpm for 10 min at room temperature. The pellet was washed in 70 % ethanol twice, air dried and resolved in 30 µl TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA).

Genomic DNA for methylation analysis was isolated by DNeasy Plant Mini Kit (Qiagen).

2.2.2.4 Southern blotting

For Southern blot analysis, 2 µg genomic DNA was digested restriction enzyme at 37 °C overnight, then transferred to Amersham Hybond™-NX membrane (GE Healthcare) after gel electrophoresis using wet capillary transfer (Lichtenstein et al. 1990). DNA was fixed to the membrane after UV-crosslinking by using Stratagene UV Stratalinker 2400.

The membrane was pre-hybridized for 1 hour in pre-hybridization solution (5x SSC, 3 % Blocking Reagent (Roche), 0.1 % Na-Laurylsarcosine, 0.2 % SDS, and 50 % Formamide) at 42 °C. Then the PCR produced DIG-11-dUTP (Roche) labeled DNA probe was boiled at 100 °C for 10 min, and added into pre-warmed fresh pre-hybridization solution after a short chill on ice. Hybridization was carried out in Hybrid 2000 machine (H.Saur) at 42 °C overnight. After the washing steps in order with W1 (2x SSC, 0.1 % SDS [v/v], room temperature), W2 (0.5x SSC, 0.1 % SDS [v/v], 65 °C), W3 (0.1x SSC, 0.1 % SDS [v/v], 65 °C), the hybridized membrane was immunodetected in anti-DIG-antibody (Roche) solution (1: 7600 dilution in 100 mM maleic acid, 150 mM NaCl, 3 % blocking reagent [w/v], PH 7.5). After washing three times in tween 20 buffer (0.3 % Tween 20 [v/v], 100 mM maleic acid, 150 mM NaCl, PH 7.5), the membrane was incubated in CSPD (chemiluminescent substrate for alkaline phosphatase) solution (100 mM Tris-Cl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 10 % CSPD [v/v]) for 5 min. The membrane was kept in plastic foil, and a film (CL-XPosure™ Film, Thermo Scientific) was exposed for 3 hours. The exposed film was incubated in X-ray developer solution (Calbe Chemie GmbH) for 2 min, in 1 % [v/v] Acetic acid (Roth) for 20 s, in X-ray fixer solution (Calbe Chemie GmbH) for 2 min and washed in water.

2.2.2.5 DNA methylation analysis

Bisulfite conversion was carried out by EpiTect Fast Bisulfite Conversion Kit (Qiagen) with 200 ng Genomic DNA. 2 µl from 20 µl eluted converted DNA was used for the first PCR (95 °C 3 min, (95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min) x 10 cycles, 16 °C hold). The second PCR used 2 µl product of the first PCR as template with the following program: 95 °C for 3 min, (95 °C for 30 s, 55 °C for 30 s, 72 °C for 20 s) x 40 cycles, 72 °C for 5 min, and hold at 16 °C. The PCR fragment was purified by NucleoSpin® Gel and PCR Clean-up (MACHEREY-NAGEL) and then ligated into pJET1.2 vector (Thermo Scientific). Plasmids isolated from positive clones were sequenced. Results obtained from sequencing were analyzed with the online tool CyMATE (http://www.cymate.org/) (Hetzl et al. 2007). Primers used for PCR were designed by

Kismeth (http://katahdin.mssm.edu/kismeth/primer_design.pl) (Gruntman et al. 2008) and are listed in table 11.

2.2.2.6 Estradiol induction

Seedlings grew on 1/2 MS medium plates containing 15 µg/ml hygromycin for 15 days first, and were then transferred on 1/2 MS medium plates with 15 µg/ml hygromycin and 10 ¼M 17² -Estradiol and induced for 2 days.

2.2.2.7 RNA extraction, cDNA synthesis and qRT-PCR

Total RNAs were isolated from leaves or seedlings by using either the Hot-Phenol method or TRISure^{TM} reagent (Bioline).

Hot-Phenol protocol:

Approximately 20 seedlings were collected in one 1.5 ml eppendorf tube with two small ion balls. After freezing the tubes in liquid nitrogen, the seedlings were pulverized by Retch® MM400. 1 ml pre-warmed (80 °C) Hot-Phenol buffer was added to the tube, thawed at room temperature, and then vortexed for 30 sec to homogenize material. 0.5 ml chloroform/isoamylalcohol (24:1) was added, vortexed for 30 s, and then centrifuged for 40 min at 4000 rpm. Upper phases were transferred into new tubes and kept on ice. 1 volume 4 M LiCl was added and well mixed by shaking, and the samples were left at -20 °C overnight.

Then the samples were melt at room temperature, mixed in the tube shortly by inverting, centrifuged for 20 min 16000g at 4°C. The pellet was dissolved in 450 µl DEPC-treated ddH_2O 0.1 volume 3 M NaOAc (pH 5.2) and 2 volume cold 100 % ethanol were added and the tube was well mixed and incubated at -80 °C for 30-60 min. It was centrifuged for 20 min at 16000 g at 4 °C, and the pellet was washed with 80 % cold ethanol. After air drying, the pellet was dissolved in 30 µl DEPC-treated ddH_2O.

Concentrations and purities of RNAs were measured with Nanodrop ND-1000 spectrophotometer.

TRISure reagent protocol:

Protocol of TRISure^{TM} reagent was described in product sheet (BIO-38032).

2 µg of *DNase I* digested (37 °C for 1 hour) total RNA was taken from each sample for cDNA synthesis. 1µl SuperScript® III Reverse Transcriptase (Life Technologies) was used in 20 µl cDNA synthesis reaction. Synthesized cDNA was 1:1 diluted with fresh double deionized water, then checked by qRT with GADH C-terminal and N-terminal primers, and ACTB gene. The concentrations of cDNA were finally calibrated and adjusted to 19±1 CT value of ACTB.

SYBR® GreenER™ Two-Step qRT-PCR Kit was utilized for qRT-PCR. Each 5 µl reaction contained 2.5 µl SYBR® GreenER™ mix, 0.5 µl cDNA, and 2 µl primers mix (0.5µM each). All qRT-PCR carried out with 3 biological replicates, one or two technical replicates depending on the purposes.

2.2.2.8 Protein and histone extraction, and immunodetection

Total Protein Extraction

200 mg plant material powder was homogenized with 600 µl protein extraction buffer (100 mM Tris-Cl pH 8.5, 50 mM NaCl, 50 mM EDTA, 5 % NP-40), incubated on ice for 20 min, and then centrifuged at 4 °C for 15 min at 13500 rpm. The supernatant was transferred into a fresh tube. The required amount was taken and mixed with 2x Laemmli buffer (20 %Glycerin, 4 % (w/v) SDS, 1 % (w/v) DTT (Dithiothreitol), 150 mM Tris•Cl pH 6.8, 50 mM EDTA pH 8.0, 0.04 %bromophenol blue (BPB)), heated at 95 °C for 5 min, and then loaded to the SDS gel.

Histone Extraction

Histone extraction was done using the previous described method (Lu et al. 2011a). 100 mg material was ground to fine powder with liquid nitrogen, and then resuspended in 3 × SDS loading buffer. The mixture was sonicated in an ice water bath for 5 min and boiled

at 95 °C for 5-10 min. Supernatants were moved to a fresh tube after centrifugation for 10 min at 15,000 g and then used for western blot.

Western blot

Protein supernatants were loaded to the SDS gel and ran for 1 h 15 min at 115 volts. After 15 min equilibration in electronic transferring buffer (192 mM Glycin, 25 mM Tris, 20 % Methanol, 0.05 % SDS, pH 8.3), the gel was sandwiched above PVDF membrane between blotting papers, then transferred for 45 min with 5 mA/cm^2. The membrane was removed from the blotting hardware and blocked in 7 % milk buffer (7 % milk powder in 1x TBST (diluted from 10x TBST: 0.25 M Tris•Cl pH 7.4, 1.5 M NaCl, 30 mMKCl, 0.5 % Tween-20)) for 1 hour at room temperature or overnight at 4 °C. First antibody (1:1000 dilution) was added to 7 % milk buffer and incubated with the membrane for 1 hour at room temperature. The membrane was washed three times each 5 min with 1 X TBST, and then incubated in second antibody (1:2000 dilution in 7 % milk buffer) solution for 1 hour at room temperature. The membrane was washed three times each 5 min with 1 X TBST again. The SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific) mixture was prepared and the membrane with the substrate was incubated between plastic membranes for 5 min. The PVDF membrane was then packed between fresh plastic membranes and exposed to a film for 10 s or sufficient time at room temperature in order to develop.

2.2.3 X-ray irradiation treatment

A.thaliana wild type (WT) Col-0 and mutants seedlings were grown for 2 weeks on half MS medium plates, and then were treated with 80 Gy X-ray (150 kV, 12 mA, 5.5 FOC, 88 min) at Julius Kühn-Institut (JKI) in Quedlinburg. Three hours after X-ray irradiation, seedlings were harvested and frozen in liquid nitrogen for further experiments.

2.2.4 RNA-seq and bioinformatics analysis

2.2.4.1 RNA-seq

Genomic DNA in 20 µg total RNA was removed by either being incubated with 10 units *DNase* I (Thermo scientific) at 37 °C for 30 min or RNeasy Plant Mini Kit (Qiagen) supplied on column *DNase* I digestion. The depletion of genomic DNA was further confirmed by PCR [95 °C 3 min; (95 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s) X 35 cycles, 72 °C for 5 min, 16 °C store] using primers Gapc1 (5‘ CACTTGAAGGGTGGTGCCAAG 3’) and Gapc2 (5‘ CCTGTTGTCGCCAACGAAGTC 3’) and 1 µl treated RNA as template. The concentrations of RNA were measured again after DNase I digestion. The RNA of all samples posessed the qualities that 260/280 > 1.8, 260/230 > 2.0.

Polyadenylated RNAs were enriched from total RNA and used for library construction. Sequencing was carried out on Illumina 2000 machine with the strategy of 100 paired-end reads.

2.2.4.2 Bioinformatics analysis

Reads obtained from RNA-seq were aligned to *Arabidopsis* genome (TAIR10) by Tophat. *De novo* assembly was done by cufflink, and then Cuffmerge generated a merged transcriptome annotation. Differentially expressed genes were called by Cuffdiff finally (Trapnell et al. 2012). Scripts used in this pipeline are as following:

```
#!/bin/bash

#Map the reads for each sample to the reference genome
tophat -p 8 -G genes.gtf -o 0Gyatm-1_thout genome 0Gyatm-1_R1.fastq 0Gyatm-1_R2.fastq
tophat -p 8 -G genes.gtf -o 0Gyatm-2_thout genome 0Gyatm-2_R1.fastq 0Gyatm-2_R2.fastq;
tophat -p 8 -G genes.gtf -o 0Gyatm-3_thout genome 0Gyatm-2_R1.fastq 0Gyatm-3_R2.fastq;
tophat -p 8 -G genes.gtf -o 80Gyatm-1_thout genome 80Gyatm-1_R1.fastq 80Gyatm-1_R2.fastq;
tophat -p 8 -G genes.gtf -o 80Gyatm-2_thout genome 80Gyatm-2_R1.fastq 80Gyatm-2_R2.fastq;
tophat -p 8 -G genes.gtf -o 80Gyatm-3_thout genome 80Gyatm-3_R1.fastq 80Gyatm-3_R2.fastq;
tophat -p 8 -G genes.gtf -o 0GyWT-1_thout genome 0GyWT-1_R1.fastq 0GyWT-1_R2.fastq;
```

```
tophat -p 8 -G genes.gtf -o 0GyWT-2_thout genome 0GyWT-2_R1.fastq 0GyWT-2_R2.fastq;
tophat -p 8 -G genes.gtf -o 0GyWT-3_thout genome 0GyWT-3_R1.fastq 0GyWT-3_R2.fastq;
tophat -p 8 -G genes.gtf -o 80GyWT-1_thout genome 80GyWT-1_R1.fastq 80GyWT-1_R2.fastq;
tophat -p 8 -G genes.gtf -o 80GyWT-2_thout genome 80GyWT-2_R1.fastq 80GyWT-2_R2.fastq;
tophat -p 8 -G genes.gtf -o 80GyWT-3_thout genome 80GyWT-3_R1.fastq 80GyWT-3_R2.fastq;

#Assemble transcripts for each sample
cufflinks -p 8 -u -o 0Gyatm-1_clout 0Gyatm-1_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 0Gyatm-2_clout 0Gyatm-2_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 0Gyatm-3_clout 0Gyatm-3_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80Gyatm-1_clout 80Gyatm-1_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80Gyatm-2_clout 80Gyatm-2_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80Gyatm-3_clout 80Gyatm-3_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 0GyWT-1_clout 0GyWT-1_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 0GyWT-2_clout 0GyWT-2_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 0GyWT-3_clout 0GyWT-3_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80GyWT-1_clout 80GyWT-1_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80GyWT-2_clout 80GyWT-2_thout/accepted_hits.bam;
cufflinks -p 8 -u -o 80GyWT-3_clout 80GyWT-3_thout/accepted_hits.bam;

#Run cuffmerge on all assemblies to creat a single merged transcriptome annotation
cuffmerge -g genes.gtf -s genome.fa -p 8 assemblies.txt

#Run Cuffdiff by using the merged transcriptome assembly along with the BAM files from TopHat for each replicate
cuffdiff -L U_atm,I_atm,U_WT,I_WT -o diff_assembleDEG -b genome.fa -p 8 -u
merged_asm/merged.gtf ./0Gyatm-1_thout/accepted_hits.bam,./0Gyatm-2_thout/accepted_hits.bam,./0Gyatm-3_thout/a
ccepted_hits.bam ./80Gyatm-1_thout/accepted_hits.bam,./80Gyatm-2_thout/accepted_hits.bam,./80Gyatm-3_thout/acc
epted_hits.bam ./0GyWT-1_thout/accepted_hits.bam,./0GyWT-2_thout/accepted_hits.bam,./0GyWT-3_thout/accepted_h
its.bam ./80GyWT-1_thout/accepted_hits.bam,./80GyWT-2_thout/accepted_hits.bam,./80GyWT-3_thout/accepted_hits.b
am
```

For TEs special analysis, a bowtie2 (Langmead and Salzberg 2012) pipeline was used. Reads were first mapped to TEs specific dataset (TAIR10_transposable_elements), counts were calculated by grep-awk function, and edgeR package (Robinson et al. 2010) was used for generated differentially regulated TEs. Scripts are present as:

```
#!/bin/bash
#this script is for bowtie2 mapping

bowtie2 -p 8 -x TAIR_TE_up100 -1 0Gyatm-1_R1.fastq -2 0Gyatm-1_R2.fastq -S
0Gyatm-1_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 0Gyatm-2_R1.fastq -2 0Gyatm-2_R2.fastq -S
```

```
0Gyatm-2_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 0Gyatm-3_R1.fastq -2 0Gyatm-3_R2.fastq -S
0Gyatm-3_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80Gyatm-1_R1.fastq -2 80Gyatm-1_R2.fastq -S
80Gyatm-1_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80Gyatm-2_R1.fastq -2 80Gyatm-2_R2.fastq -S
80Gyatm-2_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80Gyatm-3_R1.fastq -2 80Gyatm-3_R2.fastq -S
80Gyatm-3_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 0GyWT-1_R1.fastq -2 0GyWT-1_R2.fastq -S
0GyWT-1_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 0GyWT-2_R1.fastq -2 0GyWT-2_R2.fastq -S
0GyWT-2_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 0GyWT-3_R1.fastq -2 0GyWT-3_R2.fastq -S
0GyWT-3_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80GyWT-1_R1.fastq -2 80GyWT-1_R2.fastq -S
80GyWT-1_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80GyWT-2_R1.fastq -2 80GyWT-2_R2.fastq -S
80GyWT-2_bowtie2_TAIR_TE_up100.sam;
bowtie2 -p 8 -x TAIR_TE_up100 -1 80GyWT-3_R1.fastq -2 80GyWT-3_R2.fastq -S
80GyWT-3_bowtie2_TAIR_TE_up100.sam;

#Counting reads: Bowtie2

#U_atm
grep '^@' 0Gyatm-1_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0Gyatm-1_bowtie2_TAIR_TE_up100.txt;
grep '^@' 0Gyatm-2_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0Gyatm-2_bowtie2_TAIR_TE_up100.txt;
grep '^@' 0Gyatm-3_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0Gyatm-3_bowtie2_TAIR_TE_up100.txt;
#I_atm
```

```
grep '^@' 80Gyatm-1_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80Gyatm-1_bowtie2_TAIR_TE_up100.txt;
grep '^@' 80Gyatm-2_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80Gyatm-2_bowtie2_TAIR_TE_up100.txt;
grep '^@' 80Gyatm-3_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80Gyatm-3_bowtie2_TAIR_TE_up100.txt;
#U_WT
grep '^@' 0GyWT-1_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0GyWT-1_bowtie2_TAIR_TE_up100.txt;
grep '^@' 0GyWT-2_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0GyWT-2_bowtie2_TAIR_TE_up100.txt;
grep '^@' 0GyWT-3_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 0GyWT-3_bowtie2_TAIR_TE_up100.txt;
#I_WT
grep '^@' 80GyWT-1_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80GyWT-1_bowtie2_TAIR_TE_up100.txt;
```

```
grep '^@' 80GyWT-2_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80GyWT-2_bowtie2_TAIR_TE_up100.txt;
grep '^@' 80GyWT-3_bowtie2_TAIR_TE_up100.sam -v |
cut -f3 |
grep "*" -v |
sed "s/\.[0-9]\+//g" |
sort | uniq -c |
awk '{print $2 "\t" $1}' > 80GyWT-3_bowtie2_TAIR_TE_up100.txt;
```

3 Results

3.1 Generating DSBs by meganucleases

Reactivation of silent transposable element by chromosome damage has been known for decades (McClintock 1950). But many questions are still waiting for answers. Are all functional TEs in the genome activated by chromosome breakage, by random only few of them, or only certain classes? Is the epigenetic status of genomic DNA and TEs altered in the neighborhood of a chromosome break? In order to answer these questions, in this part of the work plants were generated in which one can induce DNA DSBs at defined positions in the genome, in the close vicinity of a transposable element.

3.1.1 Generation of *Arabidopsis* DSB reporter plants

In order to create plants in which DSBs can be induced in a controlled way, a two component system was designed: a first T-DNA vector was constructed that contains a cassette for the inducible expression of a meganuclease, and a second T-DNA vector which contains the recognition site for that nuclease in a *GUS* reporter gene, flanked by a *Ds* transposable element.

The estradiol-inducible *OlexA-46* promoter was used to control the expression of the downstream meganucleases (Figure 8, top (I)), *I-Sce I* (original CDS), *ISI-opA* (optimized CDS for I-Sce I protein in *Arabidopsis*), or the zinc-finger nucleases *ZFN3* or *QQR*. The *OlexA-46* promoter, which is made of eight copies of the LexA operator and a 35S minimal promoter, functions by the stimulation of an estradiol-activated XVE factor. The chimeric XVE factor consists of a bacterial repressor LexA (X), VP16 (V) which is the acidic transactivating domain, and an estrogen receptor (E) derived from humans (Zuo et al. 2000). As a reporter for DNA damage the meganuclease cleavage site was constructed into the *GUS* CDS so that a frameshift is generated (Puchta 1999b, Tovkach et al. 2009). After induced cleavage at the meganuclease target site, the DSB will be repaired by the error prone NHEJ pathway, resulting in the occasional restoration

of the *GUS* CDS. These cells will stain blue upon incubation with the substrate X-gluc.

A *miniDs* was inserted into the T-DNA as a potential marker for DSB (repair) induced epigenetic changes (Figure 8). Constructs generated in this part of work are: pZW-O1 (I-Sce I target site), pZW-O2 (ZFN3 target site), and pZW-O3 (QQR target site) containing recognition sites of nucleases, pZW-O4 (*I-Sce I* CDS), pZW-O5 (*ISI-opA* CDS), pZW-O6 (*ZFN3* CDS), and pZW-O7 (*QQR* CDS) containing CDSs of nucleases.

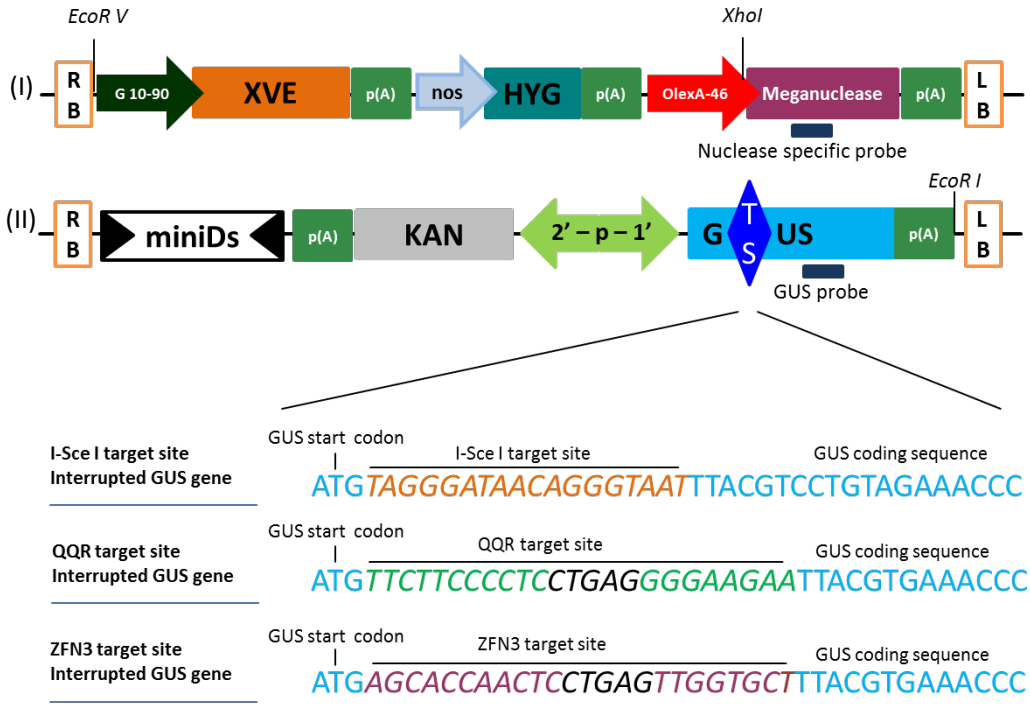


Figure 8: Structure of binary vectors and various target sites interrupted GUS CDSs.

Top: binary vectors. (I) Vector containing CDS of meganuclease. G10-90 promoter controlled XVE-element cooperates with OlexA-46 promoter to regulate expression of meganucleases. CDSs of *I-Sce I*, *ISI-opA, QQR, and ZFN3* are placed after OlexA-46 promoter. HYG: Hygromycin resistant gene. (II) Vector containing *miniDs* element and target-site interrupted *GUS* CDS. A *miniDs* element is designed for investigating the impact of vicinity break to TEs. KAN: Kanamycin resistant gene. 2'-p-1' bi-directional promoter controls both *GUS* gene and *KAN* gene. Dark blue frames represent the locations of probes for Southern blot. P(A): terminator. LB: left boder of T-DNA. RB: right boder of T-DNA.

Bottom: *GUS* CDS interrupted by various cleavage sites. All target sites are inserted following start codon. *I-Sce I* and *ISI-opA* encoded I-Sce I nuclease can recognize an 18 bp sequence (italic and dark orange). QQR (italic and green) and ZFN3 (italic and dark red) recognize 24 bp cleavage sites, respectively. QQR and ZFN3 recognition sites all contain a *Dde* I site (CTNAG) in between.

The T-DNA vectors depicted in Figure 8 were transformed into *Arabidopsis* via *Agrobacterium* mediated infection. Nuclease CDSs containing transgenic plants XVE-ISI-opA (*ISI-opA* CDS), XVE-I-Sce I (*I-Sce I* CDS), XVE-ZFN3 (*ZFN3* CDS), XVE-QQR (*QQR* CDS) were selected on half MS plates with the presence of 15 mg/L Hygromycin B. Corresponding targets and *miniDs* containing plants MIT (*miniDs* and I-Sce I/ISI-opA target site), MZT (*miniDs* and ZFN3 target site), and MQT (*miniDs* and QQR target site) were selected on half MS plates with the presence of 50 mg/L Kanamycin. Genomic DNAs were isolated from antibiotic resistant plants, and then used for genome PCR to identify transgenic plants. PCR detected transgenic positive plants were further confirmed by Southern blot. Multiple plants of each construct were identified.

Southern blot results were shown in Figure 9. Signals of positive controls and clean lanes of WT controls indicate probes specifically hybridized to target fragments. According to the possible minimum length of each target fragment, insertion pattern could be read from film (XVE-ISI-opA: 5. kb, XVE-I-Sce I: 1.5 kb, XVE-ZFN3: 1.7 kb, XVE-QQR: 5.6 kb; MIT, MZT, and MQT: 5.1 kb). Multiple insertions of XVE-ISI-opA containing T-DNA were observed in independent transgenic plants (Figure 9A). *I-Sce I* CDS containing plants were found to have two to three T-DNA insertions in transgenic offspring (Figure 9B). More than two T-DNA insertions were detected in independent transgenic plants including *ZFN3* (Figure 9C) and *QQR* (Figure 9D) CDSs. Target-sites and *miniDs* containing plants also showed different number of insertions in indepdent transgenic plants (Figure 9E). Southern blot analysis indicates target T-DNA fragments containing various meganucleases CDSs and corresponding recognition sites were successfully transformed into *Arabidopsis* plants. Confirmed transgenic plants were used for the following experiments.

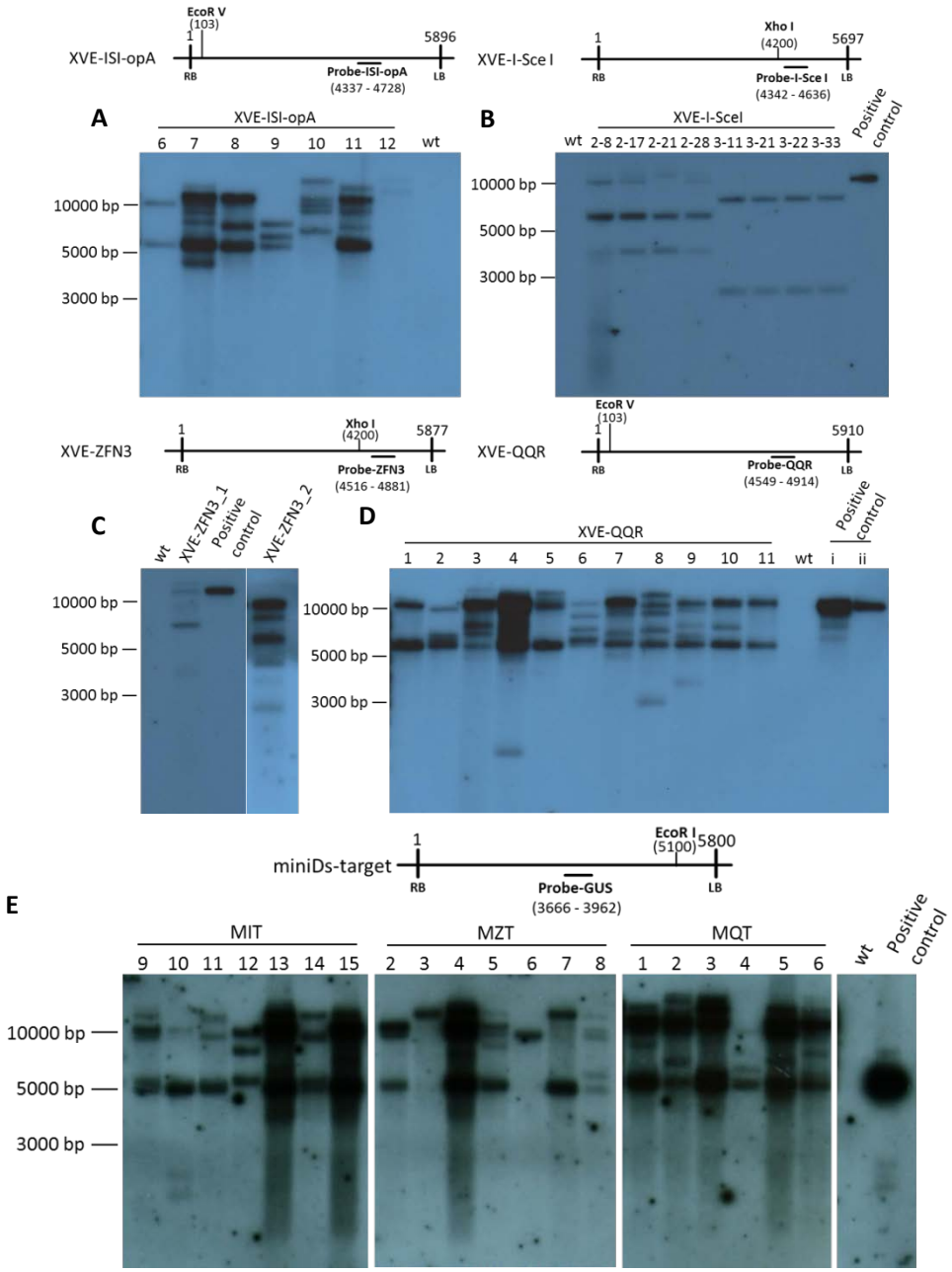


Figure 9: Southern blotting analysis of transgenic plants.

(A) XVE-ISI-opA, (B) XVE-I-Sce I, (C) XVE-ZFN3, and (D) XVE-QQR plants containing *Olex-46* promoter controlled *ISI-opA, I-Sce* I, *ZFN3*, and *QQR* CDSs, respectively. (E) MIT, MZT, and MQT plants included *miniDs* and I-Sce I and ISI-opA target site, *ZFN3* target site, and QQR target site, respectively. 2-3 µg genomic DNA from each sample was used for digestion at 37 °C overnight. Restriction enzymes used for digestion were *EcoR* V for XVE-ISI-opA and XVE-QQR plants, Xho I for XVE-I-Sce I and XVE-ZFN3 plants, and EcoR I for plants containing target sites. DIG-11-dUTP labelled probes were generated by PCR. Probes for *ISI-opA* and *I-Sce I* were amplified from *pZW-O9* and *pZW-O8*. The probe amplified from the *Fok I* sequence in *pZW-O6* was used for detecting *ZFN3* and *QQR* plants. PCR generated GUS-DIG probe was applied for all recognition sites plants, and pZW-O14 was the template for amplification. Transgenic plants with one or more T-DNA insertions were identified by Southern blot. Positive controls were corresponding plasmids used for transforming plants (XVE-I-Sce I: pZW-O4; XVE-ZFN3: pZW-O6; XVE-QQR: pZW-O7; recognition site plants: pZW-O14), and the loaded amounts were 100 pg of each sample.

3.1.2 Estradiol induced expression of ISI-opA generated DNA damage

24 h after induction with 2 µM 17-ß-estradiol, *Olex-46* promoter controlled GFP was reported to achieve its highest expressed level in 2-week old *Arabidopsis* seedlings grown on half MS plates (Zuo et al. 2000). Others observed that 10 µM estradiol is required for maximal expression levels in plants grown on half MS medium plate. However, the time period to reach the transcriptional peak is variable (Dr. Jinye Mu, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, personal communication). In this work, 2-week old seedlings containing *ISI-opA*, *I-Sce I*, *QQR*, *ZFN3* CDSs were first grown on half MS medium plates without estradiol, then transferred to fresh plates containing 10µM 17-ß-estradiol and induced for 4 h, 16 h, 48 h, and 96 h, respectively. qRT-PCR indicated that all nucleases reached high expressed levels after 4 h induction. However, *ZFN3* expression decreased slightly after 16 h induction, but still remained high throughout the sampling periods. The maximum level of I-Sce I was observed after 16 h induction. *ISI-opA* and *QQR* reached their peaks after 48 h induction (Figure 10). According to qRT-PCR results, 48 h induction was chosen for further experiments.

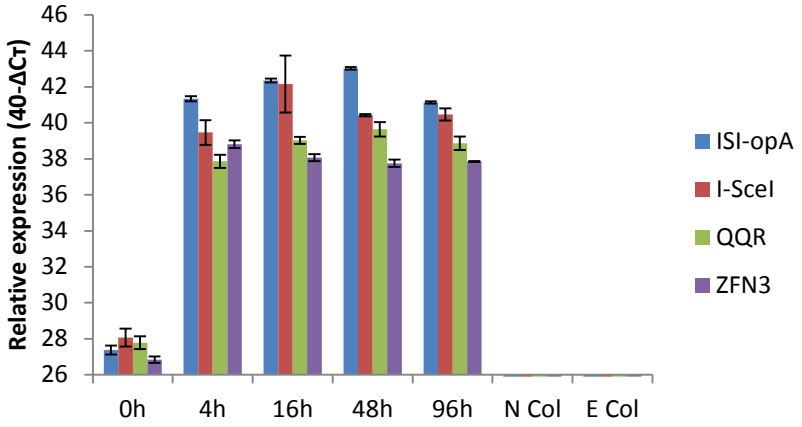


Figure 10: Detecting transcripts of nucleases after different induced periods.

Transcriptional levels were determined by qRT-PCR. Three biological replicates and two technical replicates were applied. Error bar: standard deviation. 2-week old seedlings were induced for 4 hours, 16 hours, 48 hours and 96 hours. WT plant was used as blank control. Relative quantifications were represented by 40 minus ” CB

Nucleases expressing plants with 2 T-DNA insertions were crossed with their specific target site containing heterozygous plants, respectively. Hemizygous hybrids containing both nucleases and target site T-DNAs were selected from the progeny by growing on half MS medium plates containing Kanamycin (50 mg/L) and Hygromycin B (15 mg/L). 2-week old resistant seedlings were then gently transferred to fresh half MS medium plates with 10 µM estradiol and incubated for 48 hours under normal long day conditions (Figure 11). Induced seedlings were used for further GUS assay and molecular analysis.

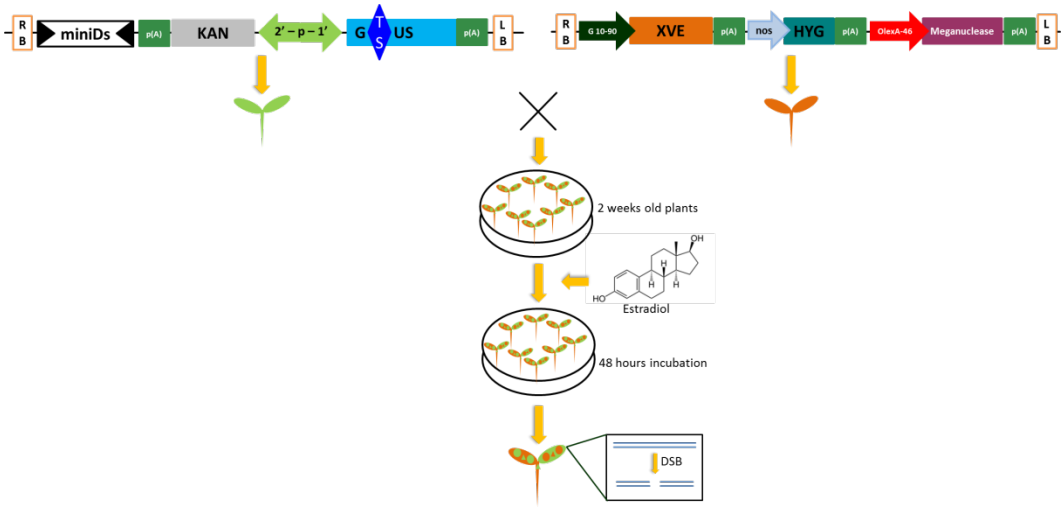


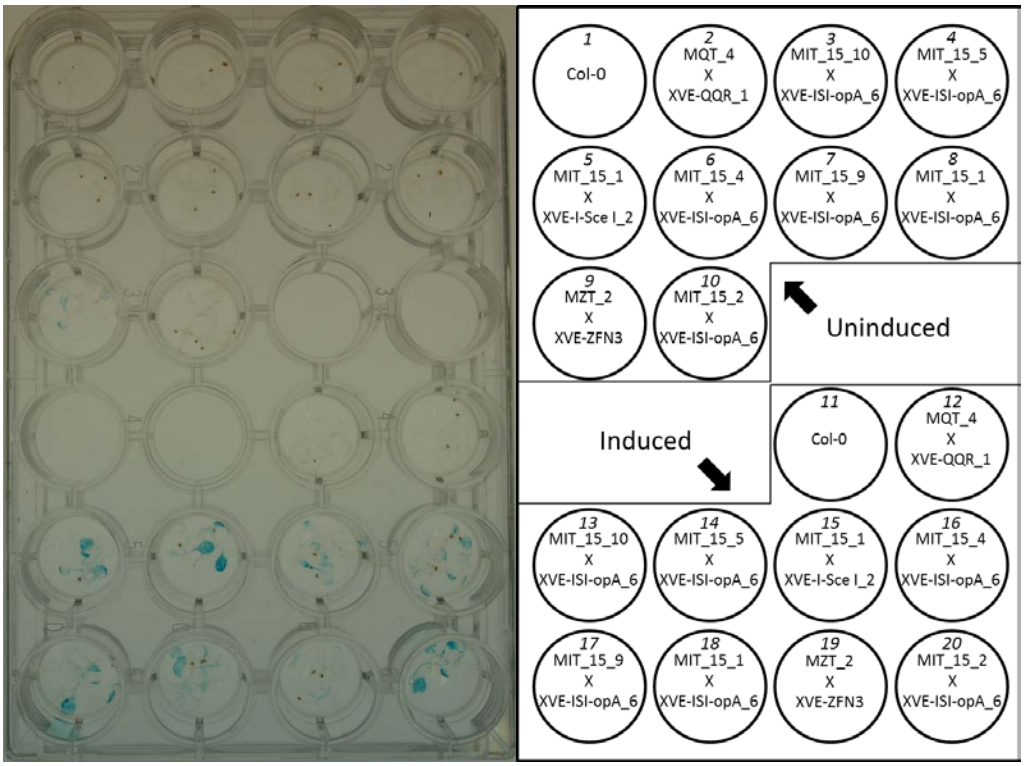
Figure 11: Illustration of estradiol induced DSBs in plants.

The *miniDs* and meganuclease target site containing plants (green plant) were crossed with their corresponding nuclease expressing plants (orange plant). The hybrid plants (green and orange mixed plant) were grown on half MS medium plate for 2 weeks, and then transferred to a fresh half MS medium plate contained 10 µM estradiol. During 48 hours induction, meganucleases were expressed, and in turn, recognized their target sites, led to DSBs.

GUS staining indicated that only I-Sce I nuclease expressing plants showed strong signals, and all ISI-opA plants had a more intense coloration in their leaves and roots than those plants expressing the unmodified original I-Sce I. This illustrated that the codon-optimized ISI-opA containing plants expressed more I-Sce I nuclease which led to higher amount of lesions in plants. Only a few blue spots appeared in ZFN3 plants, but in both induced and uninduced plants. There was no stained signal in QQR plants (Figure 12A). This indicates that neither QQR nor ZFN3 could induce enough DNA

damage or the DSBs in the GUS CDS could not be efficiently repaired in the plants.

A



B

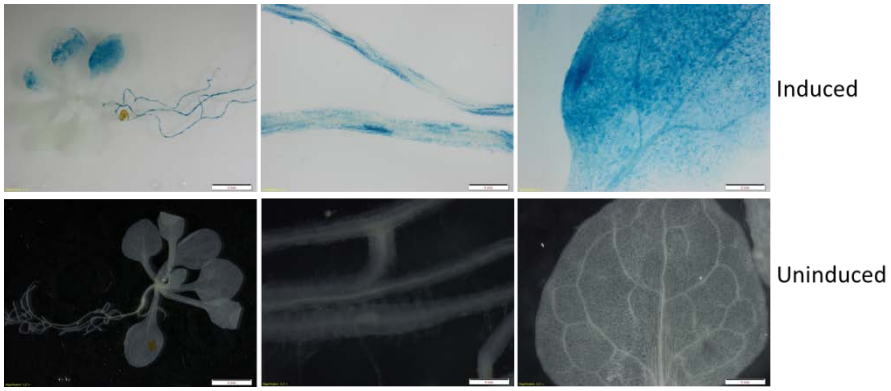


Figure 12: GUS staining of estradiol induced plants.

A) GUS staining results of uninduced (upper half) and induced (bottom half) plants. GUS signals were present specifically in estradiol induced ISI-opA and I-Sce I expressed seedlings. WT and QQR plants showed no signal after induction. B) GUS signals in MIT_15 X XVE-ISI-opA _6 plants of whole seedling, roots and leaves, after estradiol induction.

Figure 12B shows that GUS signals were present in only three of six leaves and the distribution of blue spots in leaves and roots was not even in estradiol induced ISI-opA plants. The GUS signal was absent in plants without estradiol treatment. The results indicate that DNA damage was dependent on estradiol induction in this work, but the

occurrence of lesions and their subsequent repair was not equal in plant cells.

In order to further analyze induced DNA damage and repair in plants, an *in vitro* digestion assay was applied. After DNA damage, the endogenous repairing machinery will repair the breaks. If the DSBs are repaired by NHEJ, in most cases the nuclease recognition site will be destroyed afterwards. Accordingly, after meganuclease digestion of PCR products covering the recognition sites the fraction of undigested product relative to digested product will indicate the efficiency of the preceding *in vivo* meganuclease cleavage reaction.

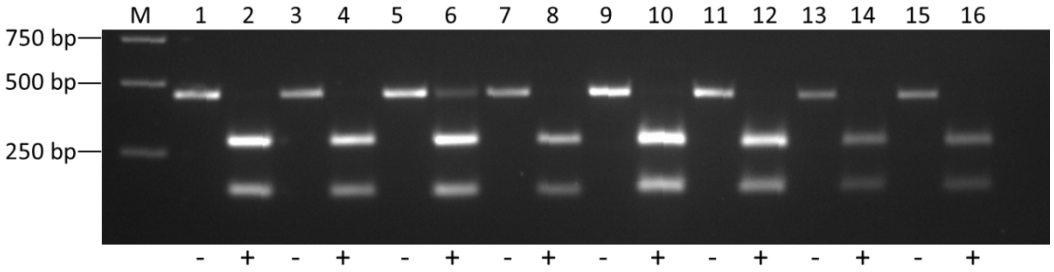


Figure 13: *In vitro* analysis of DSBR.

Without enzyme -, with enzyme +; M: Generuler 1kb ladder. 1 and 2: PCR products amplified from induced MIT X XVE-I-Sce I plants. 3 and 4: PCR products amplified from uninduced MIT X XVE-I-Sce I plants. 5 and 6: PCR products amplified from induced MIT X XVE-ISI-opA plants. 7 and 8: PCR products amplified from uninduced MIT X XVE-ISI-opA plants. 9 and 10: PCR products amplified from induced MZT X XVE-ZFN3 plants. 11 and 12: PCR products amplified from uninduced MZT X XVE-ZFN3 plants. 13 and 14: PCR products amplified from induced MQT X XVE-QQR plants. 15 and 16: PCR products amplified from uninduced MQT X XVE-QQR plants.

Genomic DNAs were isolated from induced and uninduced hemizygous F1 plants. The same amounts of PCR products from each sample were used for digestion. Commercial I-Sce I enzyme was used for digesting fragments containing the I-Sce I target site (I-Sce I and ISI-opA plants). Because QQR and ZFN3 recognition sites contain a Dde I target (Figure 8) sequence, commercial Dde I enzyme was applied for digesting *QQR* and *ZFN3* segments. Digested products were separated on 2.5 % agarose gels (Figure 13). Only ISI-opA plants showed a sharp undigested band (Figure 13, lane 6), which indicated that ISI-opA encoded I-Sce I was the most efficient nuclease to create DSBs in transgenic plants among all four nucleases. This result coincided with the GUS assay.

The intensities of all three bands derived from ISI-opA plants were measured by ImageJ. The fraction of chromosomal molecules that had been cleaved *in vivo* and repaired under destruction of the nuclease recognition site was calculated by dividing the intensity of the uncleaved 480 bp PCR band by the total intensities of all bands. The mean damage ratio in DNA from ISI-opA plants was 6.9 %.

The undigested band from MIT_15 X XVE-ISI-opA_6 plants was purified from the gel and cloned into pJET1.2 cloning vector. 23 individual successfully sequenced positive colonies showed the repairing patterns of *ISI-opA* encoded nuclease induced DNA damage in plants (Figure 14). Sequencing results indicate that deletion was the main outcome of cleavage site repair (21 of 23) while only 2 repair events resulted in insertions. The results also indicate lesions were repaired by NHEJ.

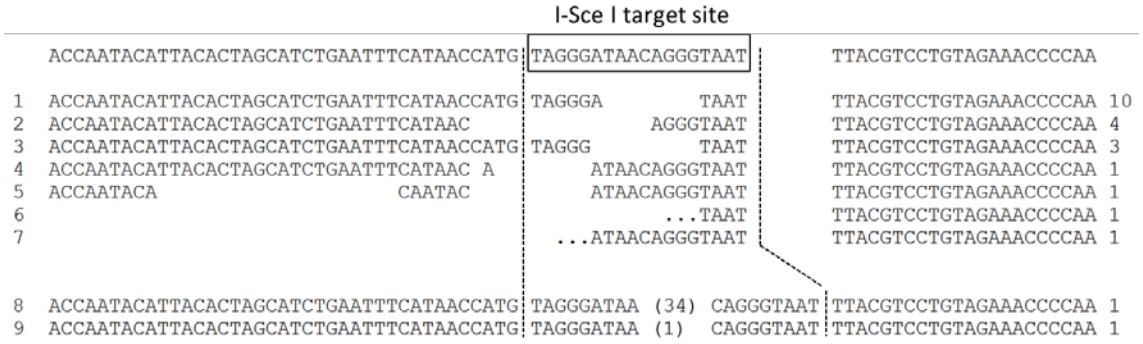


Figure 14: Repairing patterns of *ISI-opA* encoded I-Sce I induced damage in plants.

The top row shows the orginal sequence containing I-Sce I target site. Rows 1–9 show the repairing patterns after cleavage and at the right their incidence in plants. Rows 1-7 are repairing patterns with deletions. Rows 8-9 are repairing patterns with insertions. Sequence in frame is recognition site of I-Sce I.

3.1.3 Meganuclease induced DSBs did not result in transcriptional changes of detected genes

The detection of DNA repair products indicated that estradiol-induced I Sce-I expression in ISI-opA plants successfully generated DSBs. The subsequent work will determine the transcriptional alterations of candidate genes and TEs.

Mutator elements encoded mudrA and B were transcriptionally activated by UV light induced damage in maize (Questa et al. 2010). Two *Arabidopsis MUDR* homologs of *ATMUDR-a* (AT1G64260) and *ATMUDR-b* (AT1G49920) were selected. *ATCOPIA4*

(AT4G16870), *ATLANTYS2-1* (AT4G03770), *ATLANTYS2-2* (AT3G43680) and *ATLINE1-4* (AT2G01840) are methylation suppressed TEs (Lippman et al. 2003). *Mutator-like* element (AT1G36225) and two CACTA elements *CAC1* (AT2G12210) and *spm-like* (AT4G04170) were chosen for qRT-PCR analysis as well.

RAD 50 and MRE11 are signal factors of DNA damage (Paull and Lee 2005, Dupre et al. 2006). *Ku70, Ku80,* and *LIG4* are genes involved in NHEJ (Mannuss et al. 2012). These genes were chosen to be detected by qRT-PCR after induced DSBs.

DDR and regulation of TEs both are influenced by epigenetic factors, and *vice versa*. Therefore, it is obvious to determine whether epigenetic genes are regulated in response to induced DSBs. *MET1*, *CMT3* and *DRM2* encoded DNA methyltransferases maintain methylation of CG, CHG and CHH contexts in plants (Vanyushin and Ashapkin 2011). DDM1 is a chromatin-remodeling ATPase involved in cytosine methylation as well (Vanyushin and Ashapkin 2011). Demethylase ROS1 removes 5-methylcytosine from the DNA sequence (Agius et al. 2006).

Histone methylation genes *ATX1*, *ATX2, and ATXR3* are responsible for H3K4 methylation (Thorstensen et al. 2011). *KYP*, *SUVH5*, *SUVH6*, and *SUVR4* encoded proteins play roles in H3K9 methylation (Thorstensen et al. 2011). CLF and ATXR5 are able to methylate H3K27 (Makarevich et al. 2006, Jacob et al. 2009). ASH1-HOMOLOG 1 (ASHH1), ASHH2 (EFS) and ASH1-RELATED 3 (ASHR3) promote both H3K4 and H3K36 methylation (Zhao et al. 2005, Cartagena et al. 2008, Berr et al. 2009).

JMJ genes are negative regulators of histone methylation (Trewick et al. 2005). ELF6 (JMJ11), JMJ14 and JMJ15 remove methyl groups from H3K4 (Jeong et al. 2009, Chen et al. 2011). IBM1 (JMJ25) and REF6 (JMJ12) diminish methylation at H3K9 and H3K27, respectively (Saze et al. 2008, Lu et al. 2011a). The mouse JMJD5 has been reported as H3K36 demethylase (Ishimura et al. 2012), while the enzymatic function of *Arabidopsis* homolog JMJ30 has not yet been proven. Lysine Specific Demethylase 1 (LSD1) is another histone demethylase which contains no Jumonji domain (Shi et al. 2004). Its homolog of LSD1 LIKE1 removes the methyl group from H3K4 in *Arabidopsis* (Jiang et al. 2007). HDA6 is a histone deacetylase, which suppresses TEs by cooperation with METs in *Arabidopsis* (Liu et al. 2012).

Transcripts of all genes and selected TEs described above were detected by qRT-PCR in this part of the work.

ISI-opA transcript was first checked to make sure the estradiol induction was applied successfully. In induced plants, ISI-opA showed significant upregulation (Figure 15A).

ATMUDR-a (AT1G64260), *ATCOPIA4* (AT4G16870) and *spm-like element* (AT4G04170) transcriptionally decreased slightly, while *ATMUDR-b* (AT1G49920) and *CAC1* (AT2G12210) showed a tiny increase, whereas *Mutator-like* element (AT1G36225), *ATLANTYS2-1* (AT4G03770), *ATLANTYS2-2* (AT3G43680), and *ATLINE1-4* (AT2G01840) could be detected neither before nor after induction. However, none of the detected transcriptional alteration was significant (Figure 15B).

DDR genes (*Ku70, Ku80, RAD50, LIG4, MRE11,* and *SNM1*) (Figure 15C), DNA methyltransferases and demethylase (*CMT3, MET1, DRM2, DDM1* and *ROS1*) (Figure 15C), histone methylation (*ATX1, ATX2, ATXR3, ATXR5, ASHR3, ASHH1, ASHH2, KYP, SUVH5, SUVH6, SUVR4*, *and CLF*) (Figure 15D) and demethylation (*JMJ14, JMJ15, JMJ30, IBM1, REF6* and *LDL1*) (Figure 15E) related genes were detected by qRT-PCR as well. But none of them showed significant regulation after induced DNA damage in *Arabidopsis.* The only exception was *RAD50* with a slight decrease ($p = 0.019$). RAD50 is the component of the MRN complex (MRE11/RAD50/NBS1) which generates the platform for DSBR (Paull and Lee 2005, Dupre et al. 2006). Another component, MRE11, also showed tiny but insignificant reduction.

The qRT-PCR results and *in vitro* digestion assay indicated that *ISI-opA* was specifically expressed by estradiol induction and its encoded I-Sce I protein generated DSBs in plants, but the low damage ratio might be the reason why almost no selected TEs or genes showed significant transcriptional alteration after damage. It seems that the meganuclease induced damage system created in this work could not achieve the aim to investigate changes of TEs on the transcriptional level after DNA damage.

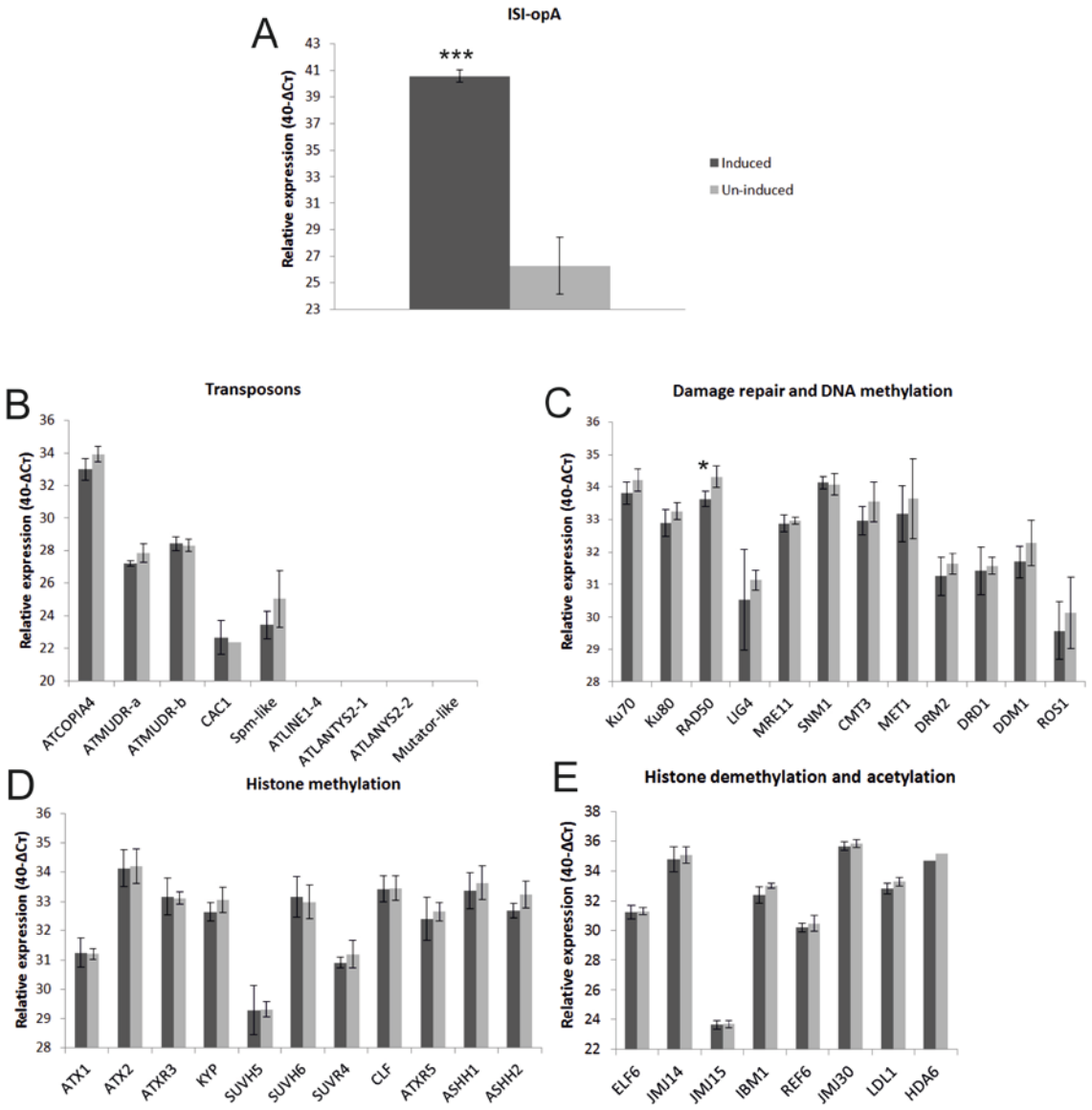


Figure 15: qRT-PCR detection of selected targets in estradiol induced ISI-opA plants.

(A) Specific expression of ISI-opA transcript in induced plants. (B) Candidates of transposons. (C) DNA damage and DNA methylation related genes. (D) Histone methyltransferase coding genes. (E) Histone demethylation and acetylation genes. 4 biological replicates were carried out for qRT-PCR. Expressions of all genes were normalized to beta actin. Error bar: standard deviation. * $P < 0.05$ (T-test). *** $P < 0.001$ (T-test).

3.2 Ionizing radiation induced transcriptomic response

The tiny damage ratio probably resulted in the undetectable transcriptional changes of genes/TEs after induced DSBs by meganuclease. Therefore, a powerful method is required to create a large amount of DSBs efficiently in plant cells. Ionizing radiation is a

good tool for this purpose. In this work, X-ray treatment was carried out in order to generate DSBs in *Arabidopsis*. ATM is a kinase recruited by DNA damage, mainly DSBs, and further controls downstream genes in response to breaks. In this part of work, *atm* mutant and WT *Arabidopsis* seedlings were irradiated with X-ray in order to investigate the transcriptional alterations of TEs after DSBs. Due to the capacity of qRT-PCR and the fact that commercial microarray does not cover all TEs, the recently developed and still developing high-throughput technology, RNA-seq, was used for investigating the whole transcriptomic changes.

3.2.1 X-ray irradiation of plants

WT and *atm* mutant seedlings were grown on half MS medium plates for two weeks (Figure 16A). In order to give only sub-lethal DNA damage to WT plants, seedlings on plates were irradiated by X-ray with a dosage of 80 Gy (X-ray machine was kindly supplied by Dr. Evelyn Klocke at the Julius Kühn-Institut (JKI) in Quedlinburg, Figure 16B and C). In previous report, ionizing radiation induced transcription already reaches a high level three hours after treatment (Ricaud et al. 2007). Since the aim of this work is to investigate the direct impact of DSBs to TEs, it is necessary to avoid the complex following sub-regulatory network. So plant materials were harvested three hours post X-ray irradiation. 20-22 seedlings were pooled into one tube, frozen in liquid nitrogen immediately and stored in minus 80 °C for further experiments. Three biological replicates were prepared for RNA-seq of irradiated and unirradiated WT and mutant plants. Materials were prepared for DNA analysis, such as DNA methylation, and protein immunodetection as well.

3.2.2 Overview of RNA-seq data

RNA-seq was carried out with at least 10 µg total RNA of each sample. Libraries were generated after poly(A) enrichment of RNAs. X-ray irradiated and unirradiated WT and *atm* mutant plants with three biological replicates, a total of 12 samples were sequenced on the Illumina HiSeq2000 machine. The sequencing types were 90 bp and 100 bp

paired-end reads, respectively. As a result, after trimming adaptors, all the samples generated more than 20 million paired-end reads, more than 350 million paired-end reads in total.

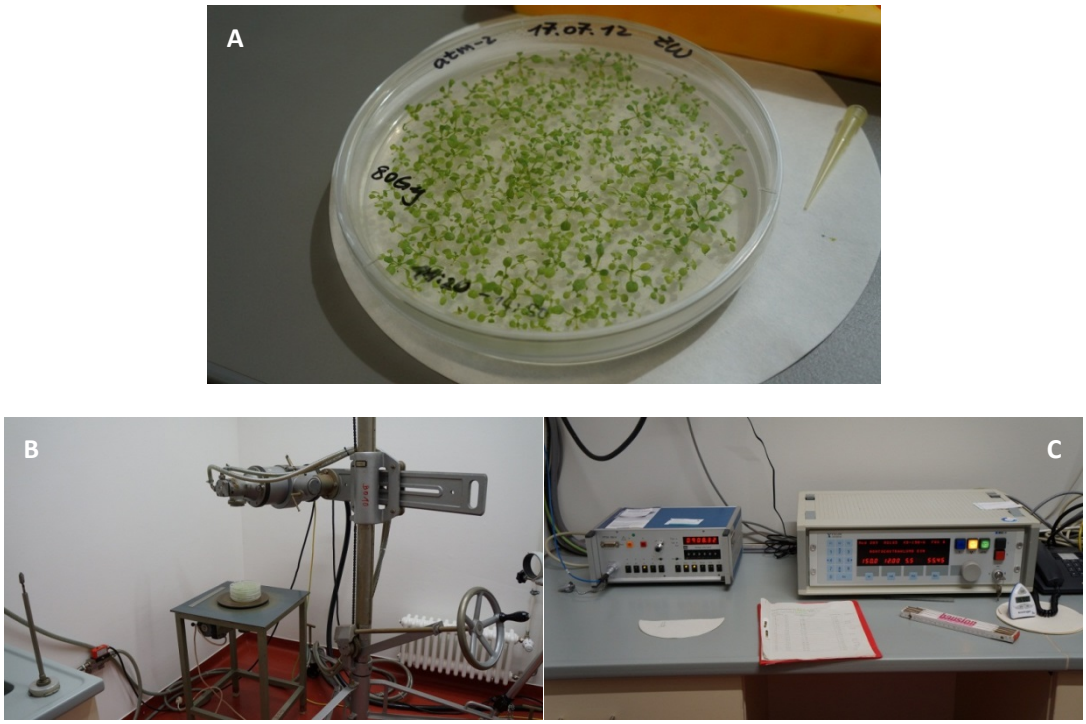


Figure 16: Plants for X-ray treatment and X-ray machine.

Arabidopsis seedlings were grown on half MS medium plates without antibiotics. 2-week old plants were treated by X-ray irradiation with 80 Gy dosages (A). Seedlings on plates were put beneath X-ray machine (B). The required dosage was adjusted by the controlling system (C).

Reads of all samples were mapped to *Arabidopsis* genomic sequence, which is the TAIR10 reference obtained from Ensembl (http://tophat.cbcb.umd.edu/igenomes.html), by Tophat, and the differential expression analysis was done by Cufflinks. The visualizations of results generated by Tophat and Cufflinks were created by the CummeRbund package (Trapnell et al. 2012). The alignment allowed up to 2 mismatches of each segment. Of the total paired-end reads obtained in X-ray untreated WT (111.6 million) and *atm* mutant (80.0 million), treated WT (93.2 million) and *atm* mutant (68.2 million) plants, 104.9 million (94.0 %), 78.1 million (97.6 %), 87.6 million (94.0 %), and 65.9 million (96.7 %) reads were aligned to the *Arabidopsis* genome, respectively (Table 12). For the scripts of Tophat and Cufflinks see 2.2.4.2.

Table 12: Ratios of mapped reads

Samples	Overall alignment rate	Replicates	Input paired-end reads (Million)	Alignment rate	multiple alignments (of aligned reads)
0Gy atm	96.7 %(78.1 million)	0Gy atm-1	22.6	97.4 %	1.3 %
		0Gy atm-2	23.4	97.8 %	1.1 %
		0Gy atm-3	34.0	97.6 %	1.1 %
0Gy WT	94.0 % (104.9 million)	0Gy WT-1	52.3	94.0 %	1.6 %
		0Gy WT-2	30.9	94.1 %	5.3 %
		0Gy WT-3	28.5	93.7 %	5.8 %
80Gy atm	96.7 % (65.9 million)	80Gy atm-1	22.4	97.2 %	1.4 %
		80Gy atm-2	22.4	96.9 %	1.5 %
		80Gy atm-3	23.4	95.9 %	1.3 %
80Gy WT	94.0 % (87.7 million)	80Gy WT-1	31.4	93.7 %	2.8 %
		80Gy WT-2	29.5	94.2 %	0.9 %
		80Gy WT-3	32.2	94.1 %	4.6 %

To estimate the relative expression level of a gene, fragments per kilobase of exon sequences per million mapped reads (FPKM) is calculated by the Tophat-Cufflinks workflow (Trapnell et al. 2012). FPKM of genes were mostly ranged between 0.003 and 316, and reached the peak at 10 (Figure 17). The patterns of distribution of FPKM were quite similar among four sample groups (Figure 17).

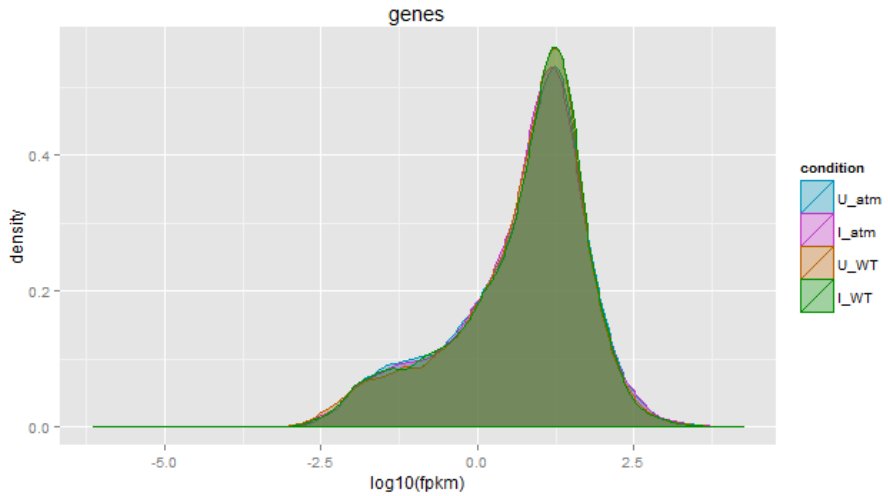


Figure 17: Density plot of individual conditions.

FPKM means fragments per kilobase of exon per million fragments mapped (Trapnell et al. 2012). The formula for FPKM is 10^9 * C / (N * L), with C is the number of mappable reads that fell onto the gene's exons, N the total number of mappable reads in the experiment and L the number of base pairs in the exon.

3.2.3 Overview of regulated transcripts in response to X-ray

Three hours post 80 Gy dose X-ray irradiation, 2336 and 1146 transcripts were significantly regulated (q-value < 0.05, q-value is adjusted p-value) in WT plants and *atm* mutants, respectively. Of these, 1315 genes in WT and 644 in the *atm* mutant were e2-fold regulated. 627 and 688 transcripts were up- and downregulated in WT plant. 345 and 299 transcripts were up- and downregulated in *atm* mutant plant (Figure 18). According to Venn diagrams, there were 528 upregulated and 582 downregulated transcripts specifically present in the WT by X-ray induction. It indicates that X-ray induced transcriptional changes are strongly ATM dependent, which is similar to gamma-irradiation induced transcriptional changes in previous report (Culligan et al. 2006, Ricaud et al. 2007). Meanwhile, 439 transcripts were only regulated in the *atm* mutant plants. It is also interesting to have a look what those genes are.

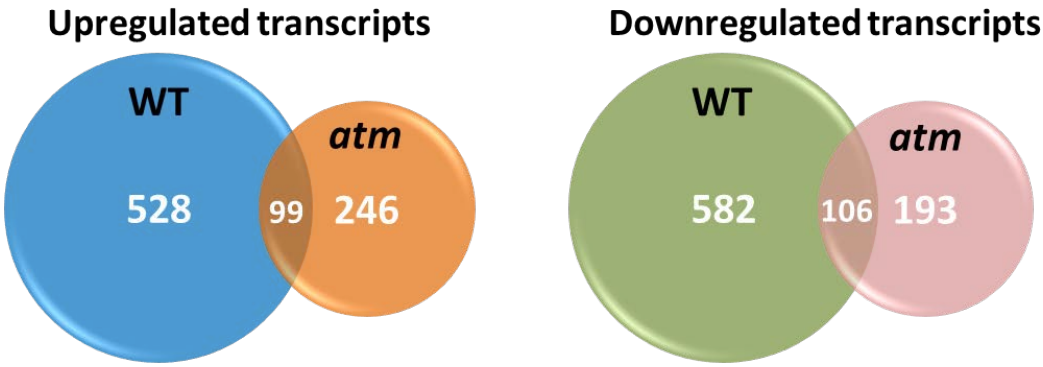


Figure 18: Venn diagrams of regulated transcripts by the induction of X-ray irradiation. (Left) In the group of upregulated transcripts, 99 transcripts shared in both WT and *atm* mutant, 528 and 246 transcripts belong to WT and mutant specifically. (Right) In the group of downregulated transcripts, 582 transcripts showed only in WT, 193 transcripts were mutant specific, 106 transcripts decreased in both WT and mutant.

In order to get an overview whether differentially expressed genes/transcripts are enriched or depleted in certain biological or functional pathways, Gene ontology enrichment analysis was carried out by the online tool – AgriGO (http://bioinfo.cau.edu.cn/agriGO/) (Du et al. 2010).

Compared to previous gamma-irradiation induced damage in *Arabidopsis* seedlings (Culligan et al. 2006, Ricaud et al. 2007), genes involved in DDR, DNA metabolism, and

ionizing radiation were highly upregualted by X-ray and in ATM-dependent manner as well. The results also revealed upregulated genes in previous papers not mentioned biological processes. Genes in indole derivative biosynthetic process were only upregulated in WT plants (Figure 19, Table S6). Tryptophan synthase genes, *CYP79B2*, *TSA1*, *and TSB2* are included in this process, which are important for downstream biosynthesis of glucosinolate that required for innate immune response (Zhao et al. 2002, Clay et al. 2009). Lipid transport proteins play an important role in defending biotic and abiotic stress (Blein et al. 2002), their coding genes specifically upregulated in *atm* mutant plants (Figure 20, Table S6) would be an interesting topic for further investigation.

Previous studies did not describe downregulated genes in details, but it is worth knowing the functions of these genes. In both WT and *atm* mutant plants, genes associated with stimulus response and cell cycle controlling were strongly downregulated (Figure 21 and 22, Table S7). However, the regulated number of genes in mutant was diminished and functional subgroups were diverse. M phase genes were particularly decreased at transcriptional level in WT plants (Figure 21, Table S7), which supports the contribution of ATM to early G2/M phase arrest (Culligan et al. 2006). The abnormal downregulation of immune genes (Figure 22, Table S7), most are TIR-NBS-LRR class genes, might lead to the death of *atm* mutant after DSBs. In *atm* mutant, a number of chitin responsive genes were downregulated by induction of X-ray (Figure 22). Most of them are *U-BOX* genes. Chitin regulates innate and adaptive immune responses in cells (Lee et al. 2008), which explains the transcriptional decrease of immune responsive genes in mutant. In the opposite, X-ray irradiation induced several chitin responsive genes upregulated in WT genes (Figure 19, Table S6). However, there was no overlap with chitin responsive genes downregulated in mutant. This probably indicates their particular functions in response to DSBs.

MapMan was used to take a closer look at regulated genes in specific pathways. It can help to check more detailed functional groups and show the expressive value of each single gene, which could not be done by GO term analysis. Three hours after irradiation, regulated putative stress induced genes were dramatically underrepresented in *atm*

mutant seedlings. 413 genes were regulated in WT, but only 188 genes in *atm* (Figure 23). Auxin response, redox, peroxidase genes were downregulated in WT, however, this was not observed in mutant plants. Smaller numbers of differentially expressed genes associated with cell wall function, protein degradation, and even the absence of heat shock proteins in mutant plants indicated their ATM-dependent pattern in response to DSBs. *WRKY* transcription factors play various roles in response to abiotic stresses (Chen et al. 2012). By induction of X-ray, reactions of *WRKY* transcription factors split into two groups: *WRKY25*, *WRKY46*, *WRKY48*, *WRKY50*, and *WRKY56* were exclusively upregulated in WT, whereas *WRKY11*, *WRKY28*, *WRKY33*, *WRKY40*, *WRKY41*, and *WRKY53* were exclusively downregulated in *atm* mutant. Another interesting group of proteins are heat shock proteins, which are only regulated in WT but not in mutant plants after irradiation. Two mainly regulated subgroups of them were small heat shock proteins and DNAJ heat shock N-terminal domain-containing proteins in WT plant.

Previous reports mentioned the epigenetic machinery is associated with DNA damge response and regulates the activities of TEs. Therefore, a closer look would be helpful to reveal functional epigenetic genes induced by DSBs. Epigenetic players include histone methylation and acetylation modifiers and DNA methylation regulators. Three hours post X-ray irradiation, general views of regulated epigenetic genes are shown in Figure 24.

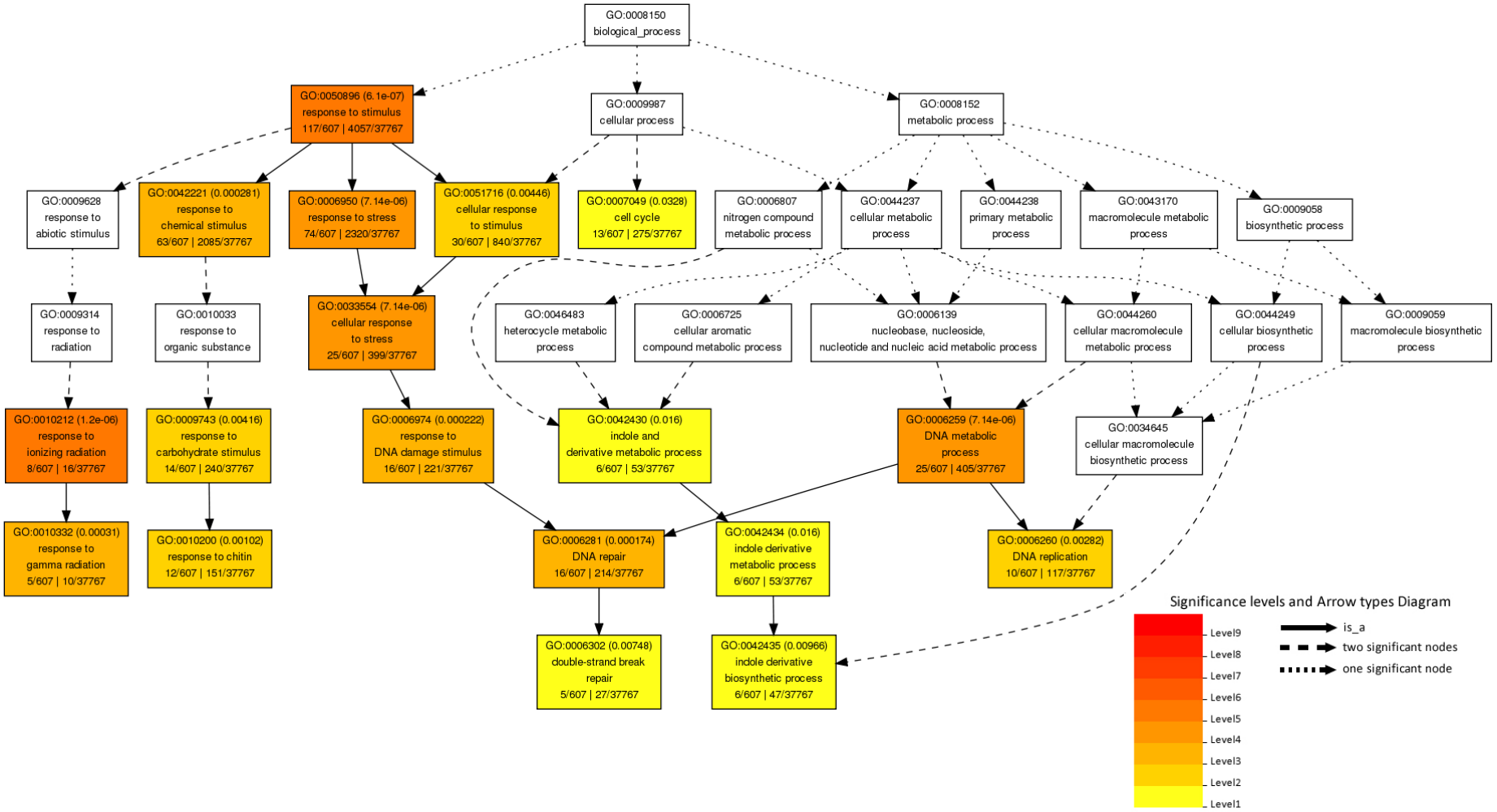


Figure 19: Partial hierachical graph of enriched Gene Ontology (GO) terms of upregulated genes in WT.

Significantly regulated functional groups are coloured (FDR < 0.05). Yellow to red represent increased significance levels.

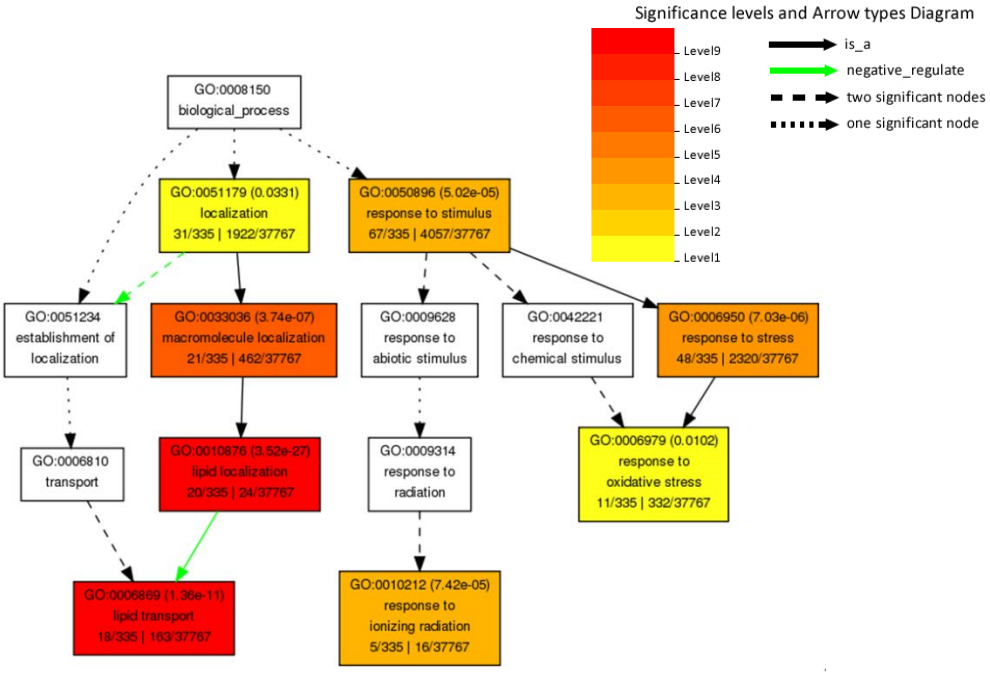


Figure 20: Partial hierachical graph of enriched GO terms of upregulated genes in *atm*.

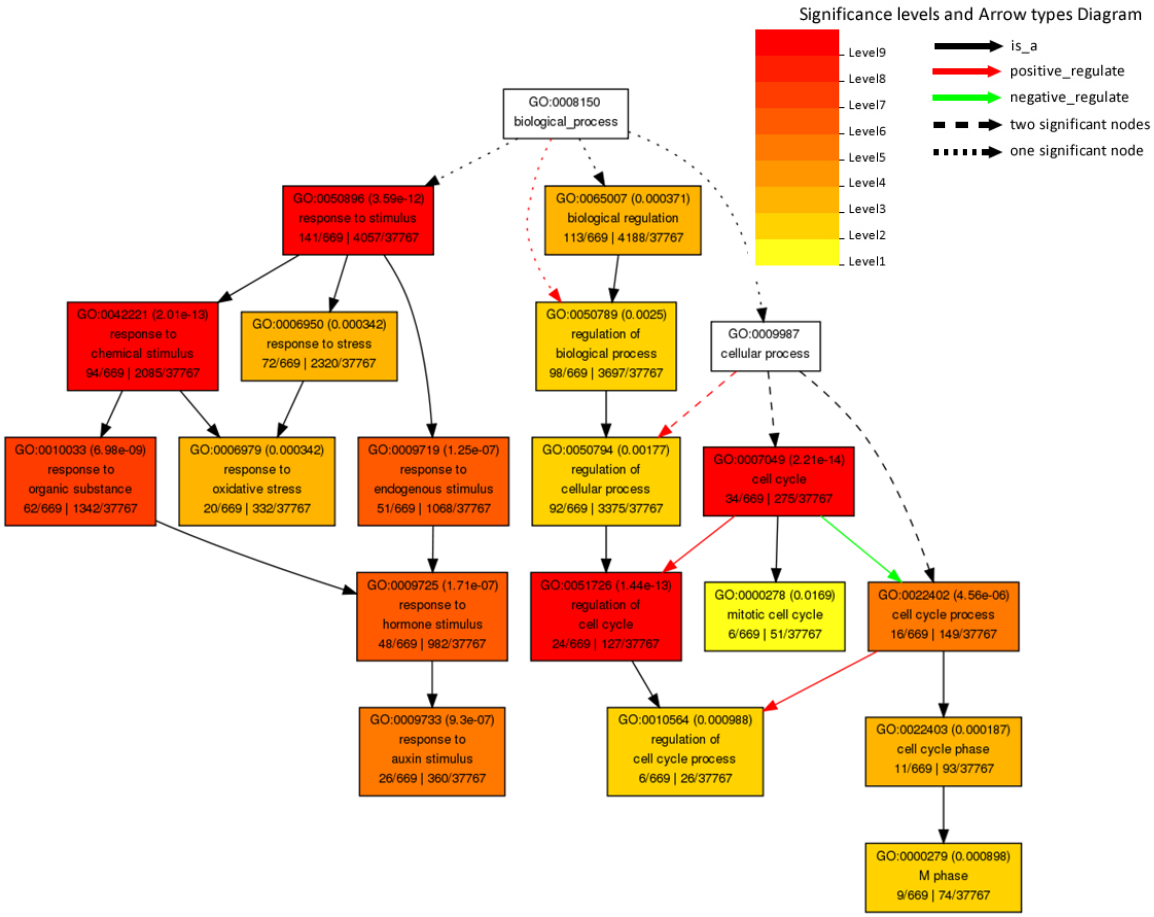


Figure 21: Partial hierachical graph of enriched GO terms of downregualted genes in WT.

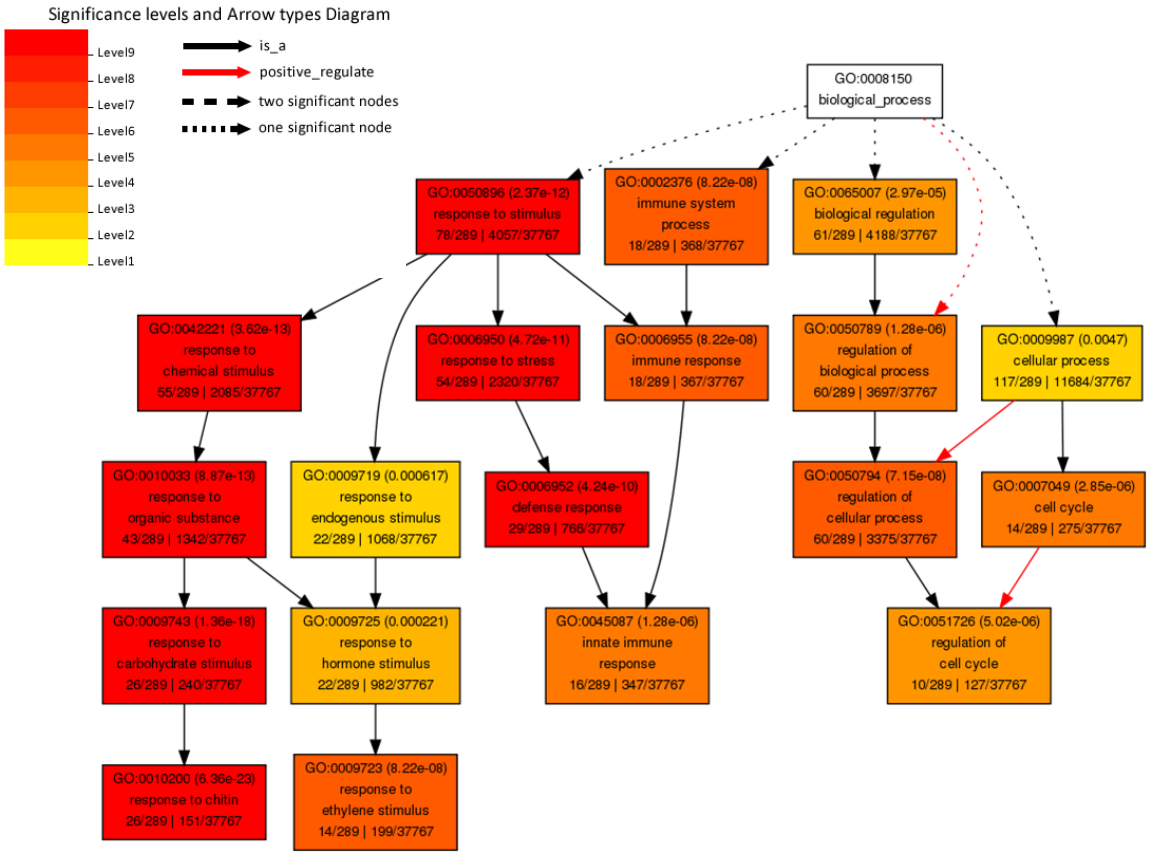


Figure 22: Partial hierachical graph of enriched GO terms of downregualted genes in *atm.*

DNA methyltransferases MET1, CMT3, and DRM2 maintain methylation at CG, CHG, CHH contexts, respectively. In WT plants, none of them showed significant transcriptional changes three hours after irradiation. *DRM1,* the homolog of *DRM2*, has only weak catalytic function of methyltransferase at CHH context. It was the only differentially expressed DNA methyltransferase with more than a 100-fold change. There are 47 predicted SET domain proteins and 21 JmjC domain-containing proteins in *Arabidopsis* (Ng et al. 2007, Lu et al. 2008). They function as methyltransferases and demethylases of histones. Only two SET domain protein coding genes *ATXR6* and *ASHR3* were differentially expressed in WT plant, which were downregulated. JMJ30 was the unique demethylase which was transcriptionally accumulated after DSBs in WT. Several genes associated with nucleosome assembly, chromatin remodeling, and histone structure were regulated in WT, such as *HMG1/2*, *HMGB6*, *Chr8*, and *Chr31*. In *atm*, only two epigenetic related genes, *DRM1* and *HMG1/2*, were differentially regulated by X-ray (Figure 24).

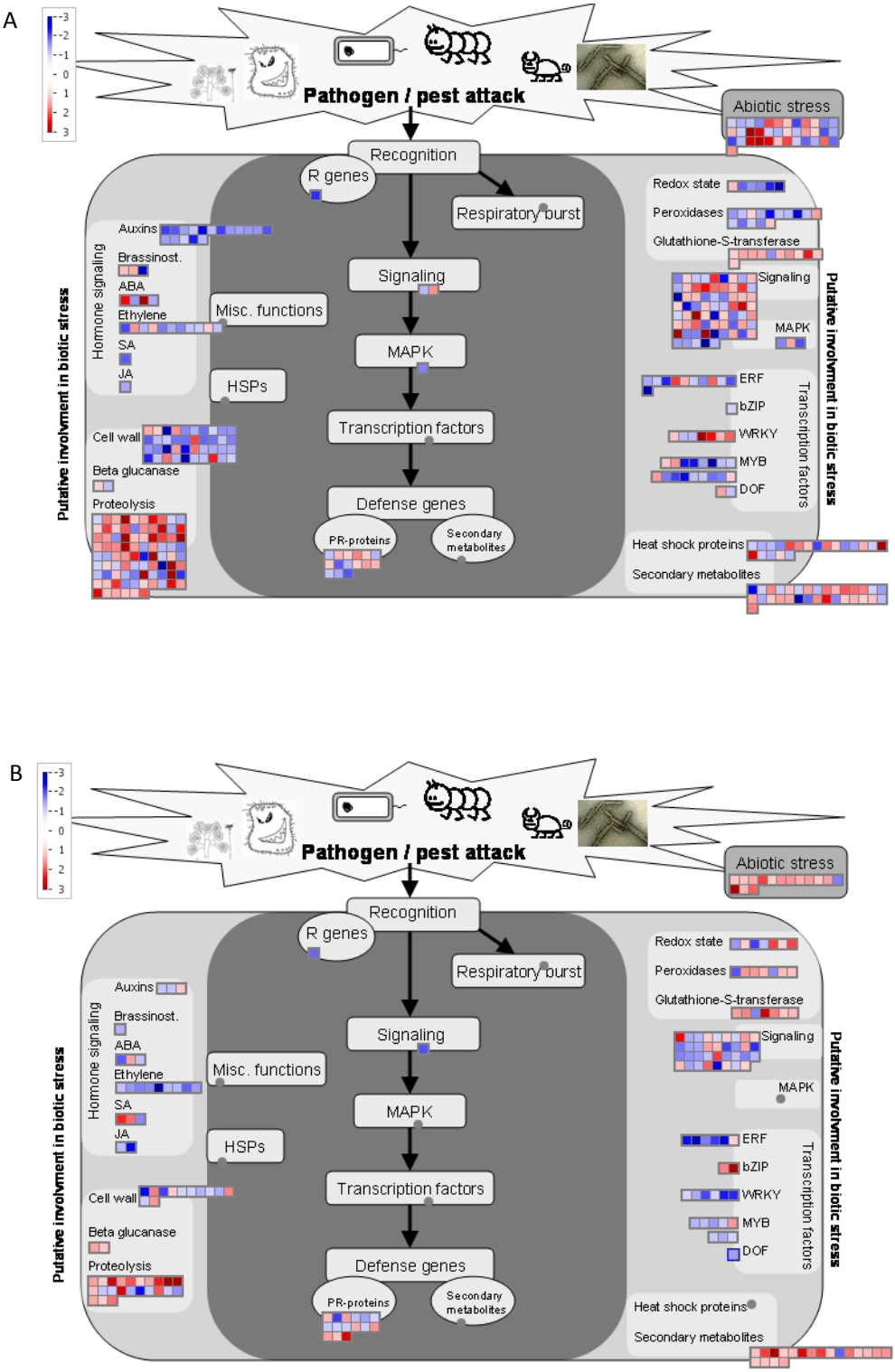


Figure 23: Stress associated genes responded to DSBs in WT (A) and *atm* mutant (B). Graphs were generated by MapMan. Red and blue in frames represent upregulation and downregulation, respectively. 413 genes were shown in WT, whereas, 188 genes were present in mutant.

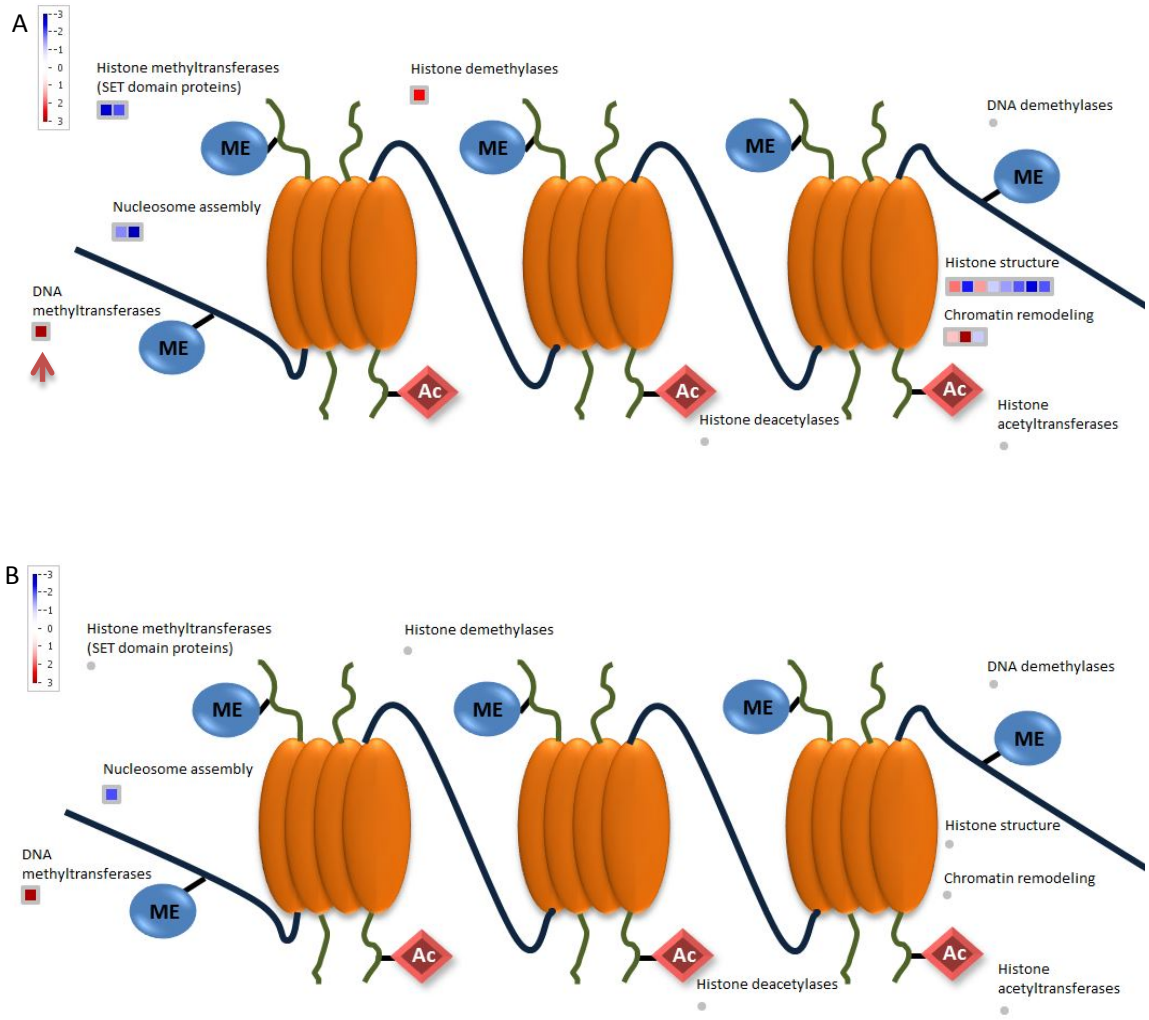


Figure 24: DSBs induced active DNA and histone methylation and demethylation, histone acetylation and deacetylation, and chromatin structure related genes in WT (A) and *atm* mutant (B). Red and blue in frames represent upregulation and downregulation, respectively. ME: methylation of cytosine or histone lysine. Ac: acetylation of histone lysine. Red arrow points out highly regulated *DRM1* gene.

As the response induced by gamma-radiation, a large number of genes were evoked to cope with DSBs three hours post X-ray treatment. More pathways in response to DNA damage are described according to RNA-seq data. Functions of genes in these pathways are worth being further investigated, especially those highly upregulated epigenetic genes, e.g. *DRM1*, *JMJ30*. Identifying their functions could improve the understanding of epigenetic regulation in DNA damage response process in plants, particularly their potential roles in regulating TEs in response to DSBs.

3.2.4 X-ray induced reactivation and suppression of TEs

For various individual transposons stimulation of transposition or reactivation from an inactive state was reported after different kinds of DNA damage in plants, animals, yeast, bacteria and humans (McClintock 1948, Peterson 1953, McClintock 1984, Rudin and Thompson 2001, Sacerdot et al. 2005, Shi et al. 2008, Questa et al. 2010). However, these observations do not allow a generalization. In order to investigate which TEs in the *Arabidopsis* genome show a transcriptional response in seedlings after X-ray treatment, RNA-seq reads were aligned to genes annotated in the TAIR10 genome release and to the transposable element sequences identified by (Buisine et al. 2008). There are 31076 putative TE-related elements in this dedicated sequence collection (Buisine et al. 2008), sequences with less than 100 bp lengths were filtered out because they can hardly be real transposons or TE relative sequences, and 23164 fragments were remained. In total, 116 and 4 elements were transcriptionally up- or downregulated in WT (Table S8) and *atm* mutant (Table S9).

Among the 116 regulated elements in WT plants, 28 were annotated as transposable element genes (Lamesch et al. 2012), which means they have potential abilities of transposition. There were 48 TE-related elements and 22 repeat elements according to repeatmasker (AFA et al. 1996-2014). The rest 18 sequences were neither TEs nor repeats (Table S8). The elements regulated in *atm* mutant contained 2 TEs/TE-related sequences, 1 repeat element (Table S9). After re-classification, 76 TEs/TEs-related elements were regulated in WT by X-ray induction. The proportions of differentially expressed genes were 4.8 % (1315 / 27416) and 2.3 % (644 / 27416) in WT and *atm* mutant (total number of genes in *Arabidopsis* is refer to latest annotation (Lamesch et al. 2012)), respectively. However, the differentially regulated TEs/TE-related sequences were only 0.32 % (76 / 23614) and 0.01 % (2 / 23614) in WT and *atm* mutant, respectively. This illustrates that TEs are, surprisingly, not as sensitive as normal genes to X-ray irradiation in *Arabidopsis*, but their response to DSBs is still ATM-dependent.

In WT, 42 and 34 TEs/TEs-related elements were up- and down-regulated, respectively (Figure 25). *DNA/MuDR* and *RC/Helitron* were two mainly regulated DNA superfamilies.

The *LTR/Copia*, *LTR/Gypsy* and *LINE/L1* were highly responsive retrotransposon families (Figure 25). Previous work always mentioned the activation of TEs after induced damage. However, interestingly, our results revealed that TEs were both transcriptional up- and downregulated by X-ray induction. Another interesting observation is that retrotransposons (53 %, 40 of 76) are more responsive to X-ray than DNA transposons (47 %, 36 of 76) in WT plants.

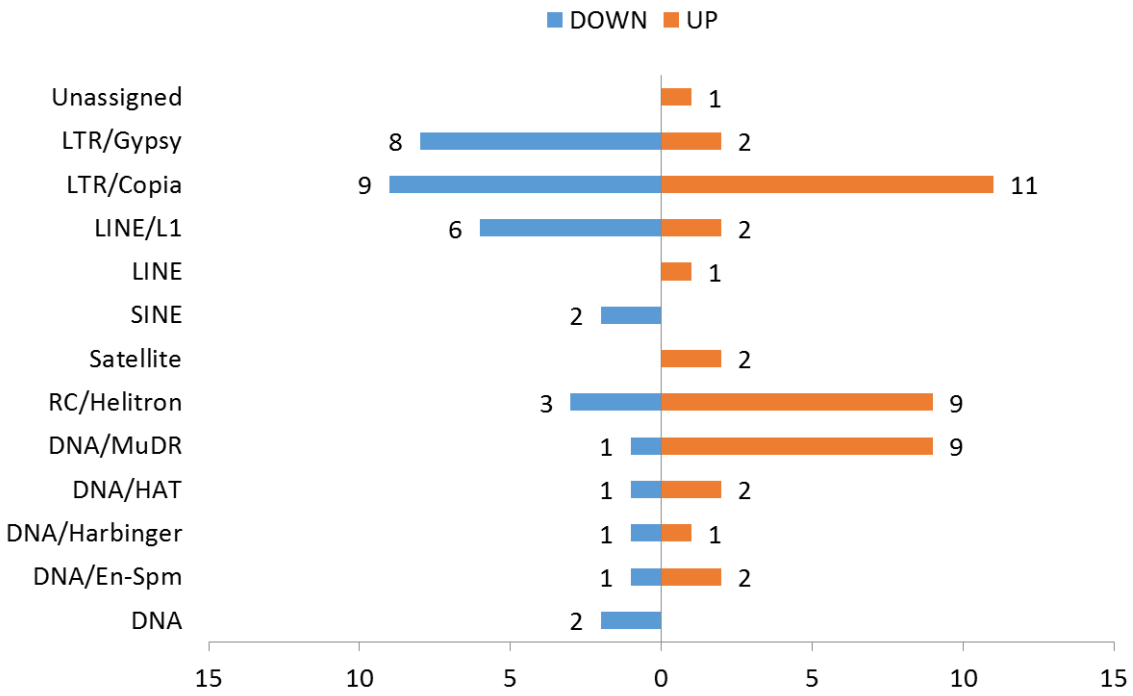


Figure 25: Regulated transposons in WT plants. TE families are listed on the left. Orange bars represent upregulated TEs, Blue bars represent downregulated TEs. Numbers of each regulated family are shown beside bars.

X-ray induced transcriptional changes of several TEs/TEs-related elements were confirmed by qRT-PCR (Figure 26). The transcripts detected by qRT-PCR showed the same trends as in RNA-seq data.

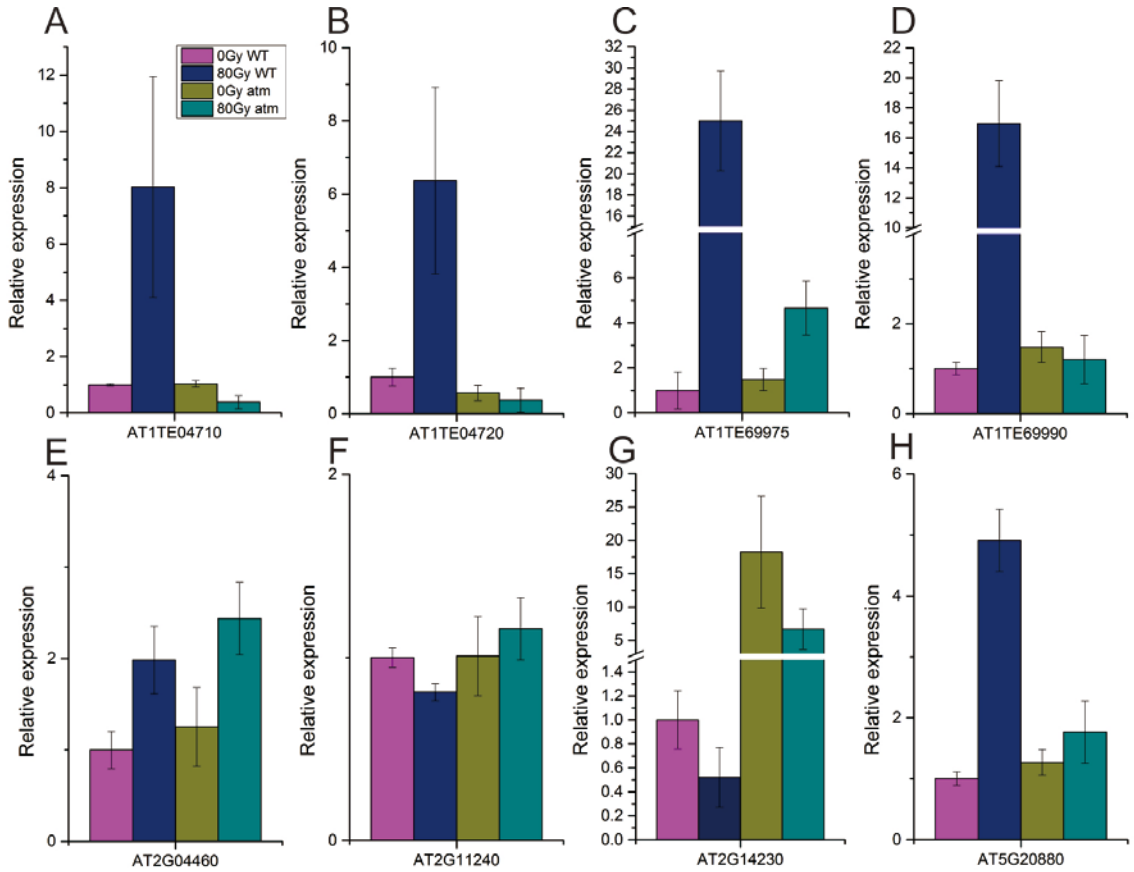


Figure 26: qRT-PCR detection of TEs/repeats.

A – D: TEs/repeats-related sequences. E – H: annotated TE genes. qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation.

3.2.5 Novel transcripts are all long non-coding RNAs

Besides annotated genes, 179 novel transcripts were detected by RNA-seq in plants. 166 of them originated from *Arabidopsis* chromosome 1 to 5, the remaining 13 were aligned to mitochondrial DNA. By checking location information of all 166 transcripts manually on PlncDB (Plant Long noncoding RNA Database) (Jin et al. 2013), 160 of the 166 transcripts were long non-coding RNAs (lncRNAs). When the sequences of remaining 6 completely novel transcripts were submitted to the Coding Potential Calculator (CPC, http://cpc.cbi.pku.edu.cn/) (Kong et al. 2007), it turned out that these 6 transcripts were also predicted to be non-coding RNAs (Table 13). In WT, 9 lncRNAs transcripts were differentially expressed, whereas, there were 8 lncRNAs in *atm* mutant plants (Table 14). Reads of each lncRNAs mapped to their genomic locations were visualized by the Integrated Genome Browser (Nicol et al. 2009). XLOC_024976

(RLFS_026432) and XLOC_024441 (At4NC060340) are two examples shown in IGB (Figure 27). There is no annotated gene at loci 4:16891269-16891936 (chromosome 4, from 16891269-16891936) and 4:14916844-14917687 (chromosome 4, from 14916844-14917687) (Figure 27). The higher and denser bars indicate the transcriptional upregulation of them in WT plants three hours after irradiation. The RNA-seq results for these two lncRNAs RLFS_026432 and At4NC060340 were verified by qRT-PCR (Figure 28), which confirmed the differential regulation in WT and mutant.

Table 13: Non-coding transcripts predicted by CPC

ID	Locus	CPC prediction
XLOC_008131	1:79171-79922	noncoding
XLOC_008133	1:8154028-8154453	noncoding
XLOC_019734	3:21971125-21972063	noncoding (weak)
XLOC_023290	4:6777265-6777672	noncoding
XLOC_024977	4:17080991-17081401	noncoding
XLOC_032251	5:8811426-8812283	noncoding

Table 14: Differentially expressed novel transcripts in WT and *atm* plants

ID	Name of lncRNA	Locus	Log2FC (WT, 80 Gy / 0 Gy)	Log2FC (*atm*, 80 Gy / 0 Gy)
XLOC_024976	RLFS_026432	4:16891269-16891936	7.22	5.11
XLOC_024441	At4NC060340	4:14916844-14917687	3.64	2.58
XLOC_006240	RLFS_006415	1:16416171-16417862	2.55	-
XLOC_008153	RLFS_007748	1:26209376-26210704	2.2	1.86
XLOC_008145	At1NC064450	1:17292412-17292998	1.92	
XLOC_027207	At5NC066580	5:17045344-17047221	1.6	2.24
XLOC_019295	RLFS_021173	3:19152296-19153335	1.27	1.81
XLOC_009679	At2NC048830	2:11688078-11689376	-1.75	-
XLOC_013763	At3NC007260	3:1966670-1967469	-2.58	-
XLOC_013612	RLFS_014451	3:872193-873509	-	-1.56
XLOC_032116	RLFS_033703	5:26146596-26147719	-	-1.87
XLOC_025146	RLFS_026616	5:1041895-1042846	-	-3.28

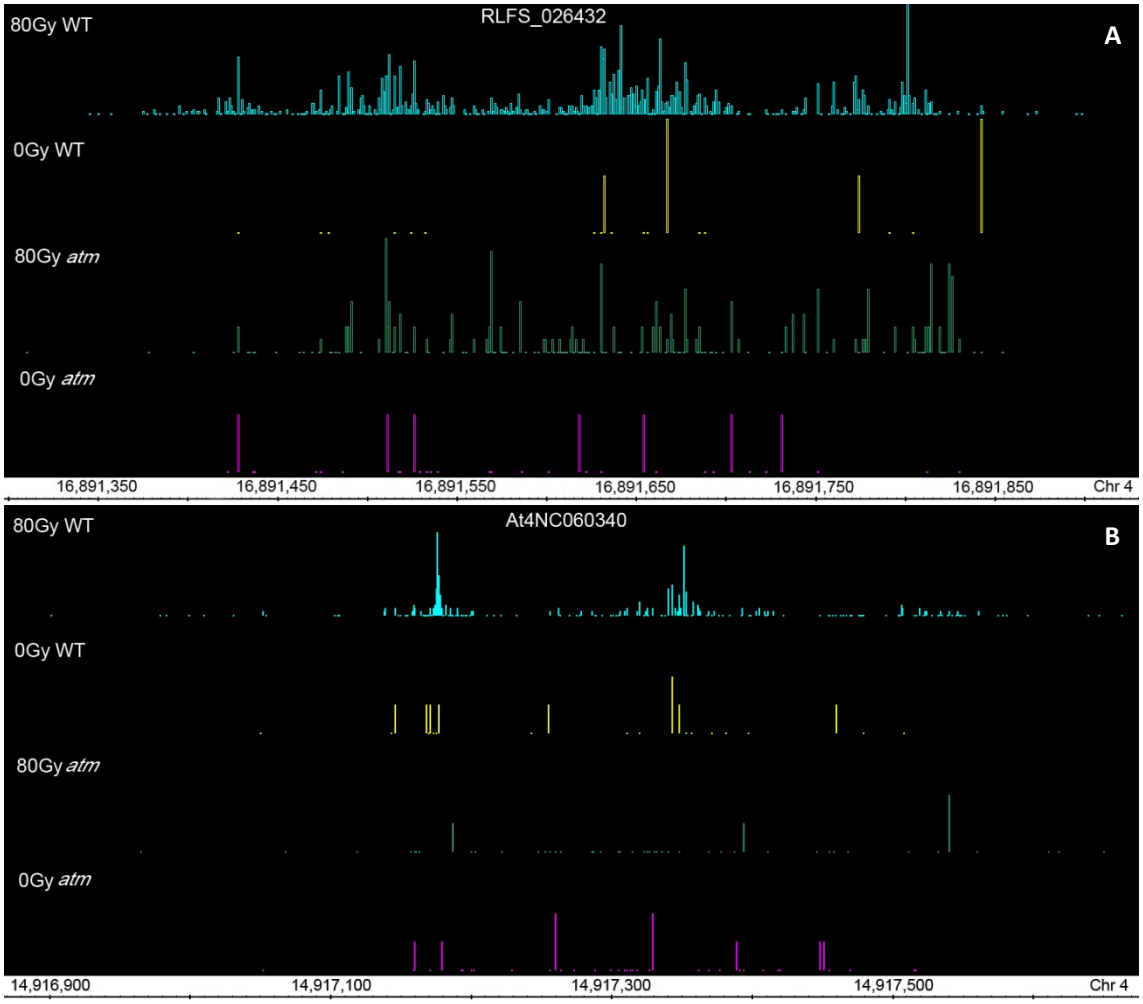


Figure 27: Visualizing RNA-seq data of RLFS_026432 (A) and At4NC060340 (B) by IGB. Reads obtained from 80Gy WT (blue), 0Gy WT (yellow), 80Gy *atm* (green), and 0Gy *atm* (purple) were mapped to loci of RLFS_026432 and At4NC060340. In WT plant, three hours after 80 Gy X-ray treatment, higher and denser bars represent more reads mapped to RLFS_026432 and At4NC060340 loci, indicate their upregulations.

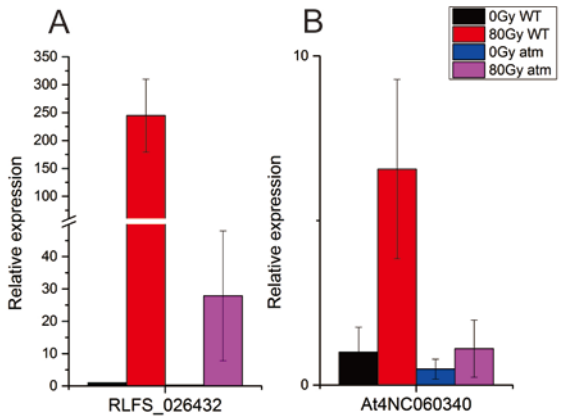


Figure 28: Detected expressions of lncRNAs RLFS_026432 and At4NC060340 by qRT-PCR. qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation.

3.2.6 ATM regulates lncRNAs in response to DSBs

After finding out TEs/TE-related sequences from the 116 regulated elements in WT, it is interesting to know more about those repeats and sequences without characterizations. By checking their locations via plant long non-coding RNA database (PLncDB) (Jin et al. 2013) or predicting their coding potentials by CPC (Kong et al. 2007), they were all found to be non-coding RNAs (ncRNAs). Intriguingly, locations of 48 TE-related sequences (Table S8) are in, overlapping or containing annotated lncRNAs regions or are predicted to be ncRNAs, and even 4 TE genes contain lncRNAs as well.

LncRNAs are a group of transcripts mainly generated by Pol II and additional by two plant specific RNA polymerases, Pol IV and Pol V (Wierzbicki 2012). LncRNAs are considered to have important functions in regulating gene expression and chromatin modifications (Kim and Sung 2012, Wierzbicki 2012, Zhang and Chen 2013). They are presumed to function mainly via guiding protein complexes to target loci depending on specific similarity or as scaffold for siRNAs, and then alter the chromatin modification or DNA methylation (Rowley et al. 2011, Xin et al. 2011, Engreitz et al. 2013), and thus eventually regulate gene expression.

Combining the ncRNAs detected by basically gene detection and TE specific analysis, three hours post X-ray irradiation, 91 and 18 lncRNAs were regulated in WT and atm mutant plants, respectively. Those lncRNAs contains 9 and 8 novel transcripts detected in WT and *atm* (Table 14). They are not annotated as any coding genes, but are lncRNAs according to checking their loci on PLncDB. Intriguingly, in WT, 73 out of 91 lncRNAs are derived from TEs/repeats associated region. In *atm*, there are 8 TEs/repeats originated lncRNAs as well (Figure 29).

The observation indicates that lncRNAs are regulated by ATM in response to X-ray induced DSBs in *Arabidopsis*. Those lncRNAs are largely transcribed from TEs/repeats sequences.

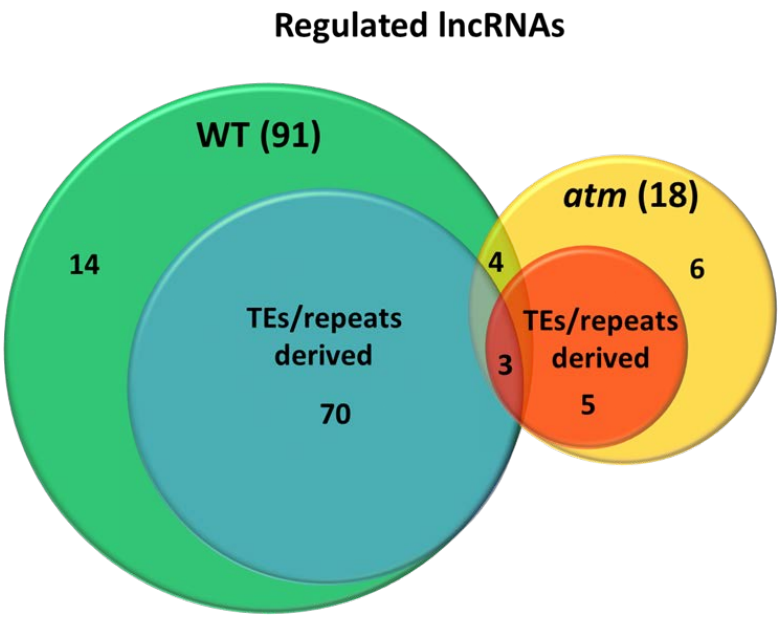


Figure 29: Venn diagram of regulated lncRNAs in WT and *atm* mutant.

91 lncRNAs induced by DSBs in WT, only 18 present in *atm* mutant. In WT, 73 out of 91 lncRNAs are derived from TEs/repeats. In *atm* mutant, 8 out of 18 lncRNAs are TEs/repeats related. There are only 7 shared lncRNAs detected in both WT and *atm* mutant by induction of X-ray, 3 are TEs/repeats original.

3.2.7 Co-regulation of lncRNA-flanking genes

Among regulated lncRNAs, 40 % in WT (Table S10) and 28 % in *atm* mutant (Table S11) have co-expressed genes. These lncRNAs and genes reacted similarly to X-ray irradiation at transcriptional level.

In WT plant, two regulated lncRNAs derived from AT1TE69975 and AT1TE69990 locate in the first and fourth introns of *WRR4* (White Rust Resistance 4) (Figure 30A), which is a defense gene. They were all upregulated three hours after X-ray treatment. In order to investigate if those two lncRNAs are independent transcripts, semi-quantitative RT-PCR was used for the detection. PCR results showed that primer pair started at *WRR4* exon

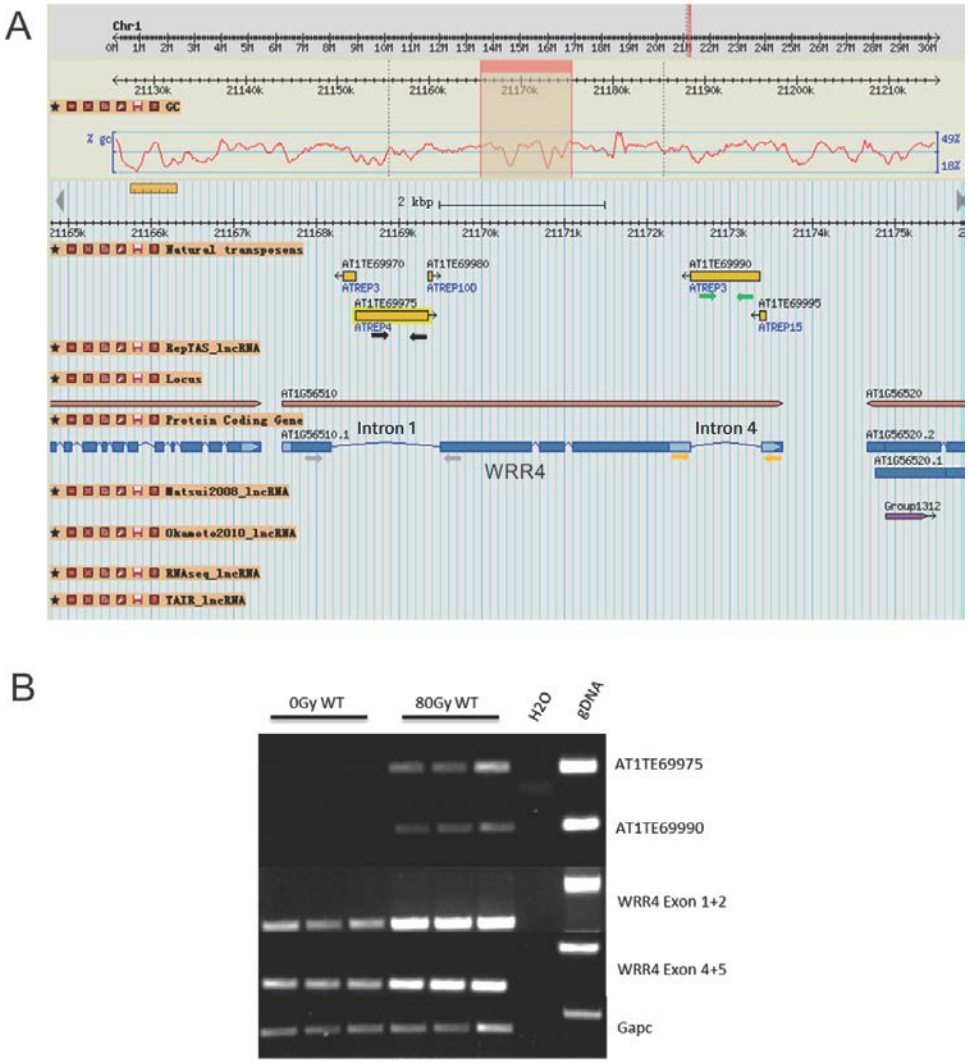


Figure 30: Illustration of AT1TE69975 and AT1TE69990 located in introns of *WRR4* gene and semi RT-PCR confirmation. (A) All information was obtained from PlncDB, protein coding genes, various lncRNAs datasets and Natural transposons were selected to be shown in graph. AT1TE69975 inserted in intron 1 and AT1TE69990 inserted in intron 4 of *WRR4*. Arrow pairs illustrate positions of primer pairs used for semi RT-PCR. Black pair: ZW_P569 and ZW_P570 for AT1TE69975. Green pair: ZW_P575 and ZW_P576 for AT1TE69990. Grey pair: ZW_P565 and ZW_P566 for amplifying the region crossing intron1 of *WRR4*. Yellow pair: ZW_P567 and ZW_P568 for amplifying the region crossing intron4 of *WRR4*. (B) Semi-quantitative RT-PCR showed that AT1TE69975, AT1TE69990 and *WRR4* were all induced by X-ray irradiation. Transcripts of AT1TE69975: black pair primers. Transcripts of AT1TE69990: green pair primers. Transcripts of WRR4 Exon 1+2: bl grey pair primers. Transcripts of WRR4 Exon 4+5: yellow pair primers. Gapc: reference gene.

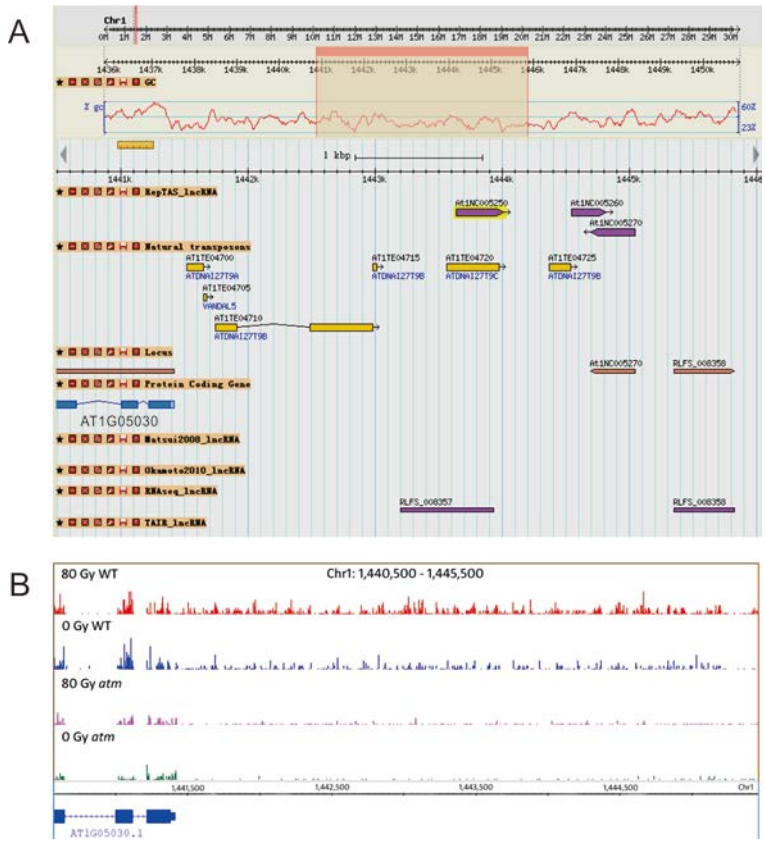


Figure 31: Illustration of AT1TE04710 and AT1TE04720 location and visualization of reads mapped to this region. (A) AT1TE04710 and AT1TE04720 (B) Visualizing RNA-seq data at region Chr1:1440500-1445500 by IGB. Reads obtained from 80Gy WT (blue), 0Gy WT (yellow), 80Gy *atm* (green), and 0Gy *atm* (purple) were mapped to Chr1:1440500-1445500 containing part of AT1G05030, and TEs/repeats fragments AT1TE04710, AT1TE04720, and AT1TE04725. In WT plant, three hours after 80 Gy X-ray treatment, higher and denser bars represent more reads mapped to upstream region of AT1G05030, indicate induced upregulation in this region and likely the existence of a long transcript containing all AT1TE04710, AT1TE04720, and AT1TE04725 elements.

1 and ended at *WRR4* exon 2 generated only one product which fits the length without intron 1. The second primer pair amplified one fragment without intron 4. Primers for AT1TE69975 and AT1TE69990 yielded PCR products with expected lengths (Figure 30B). Semi RT-PCR results confirmed AT1TE69975 and AT1TE69990 were independent transcripts of *WRR4* gene, and they were all transcriptional upregulated after DSBs.

AT1TE04710 and AT1TE04720 are lncRNAs close to each other and locate in intergenic region (Figure 31A). They were upregulated by X-ray irradiation.

AT1TE04710 and AT1TE04720 are upstream of gene AT1G05030, which was slightly upregulated after DSBs as well. AT1TE04720 overlaps with two described lncRNAs At1NC005250 and RLFS_008357 (Figure 31A). In region Chr1:1441500-1445000 containing AT1TE04710 and AT1TE04720, several other lncRNAs are also mentioned (Figure 31A). RNAseq data showed that it is very likely to be a continuously transcribed region which is upstream of AT1G05030 CDS (Figure 31B).

LncRNAs inserted between two coding genes and co-expressed with both of them were observed as well. For instance, AT4TE84220 and AT4TE84225 locate between AT4G37022 and AT4G37030 (Table S10). They were all induced by X-ray irradiation in WT plants.

Co-regulated lncRNAs and their associated or near-by genes existed in both WT and *atm* mutant. Furthermore, 40 % and 28 % of lncRNAs in WT and in *atm* mutant co-expressed with near-by genes suggests the co-regulation by X-ray induction is not coincidence. The decreased percentage of co-expression in *atm* mutant probably indicated the potential function of ATM in regulation.

3.2.8 DNA methylation of selected loci and immunodetection of Histone modifications after X-ray irradiation

RNA-seq revealed that only a few epigenetic genes were differentially expressed three hours after irradiation. Differential expression of DNA methyltransferases genes, *MET1, CMT3, DRM1* and *DMR2,* and histone methylation related genes, *JMJ30, ASHR3* and *ATXR6* were confirmed by qRT-PCR (Figure 32). Three hours post-X-ray irradiation, *MET1* was slightly downregulated in WT plants. There were no expression differences occurred with *DRM2*. Interestingly, *DRM1*, which only has weak enzymatic function at CHH methylation, was strongly upregulated by X-ray in the WT plant, whereas the increase was much lower in *atm* mutant plants. *CMT3* was decreased 2-fold by detection of qRT-PCR. *JMJ30* was upregulated with an approximate 15-fold change 3 hours after X-ray induction in the WT plants. *ASHR3* and *ATXR6*, were strongly downregulated, and showed more than 4-fold decrease.

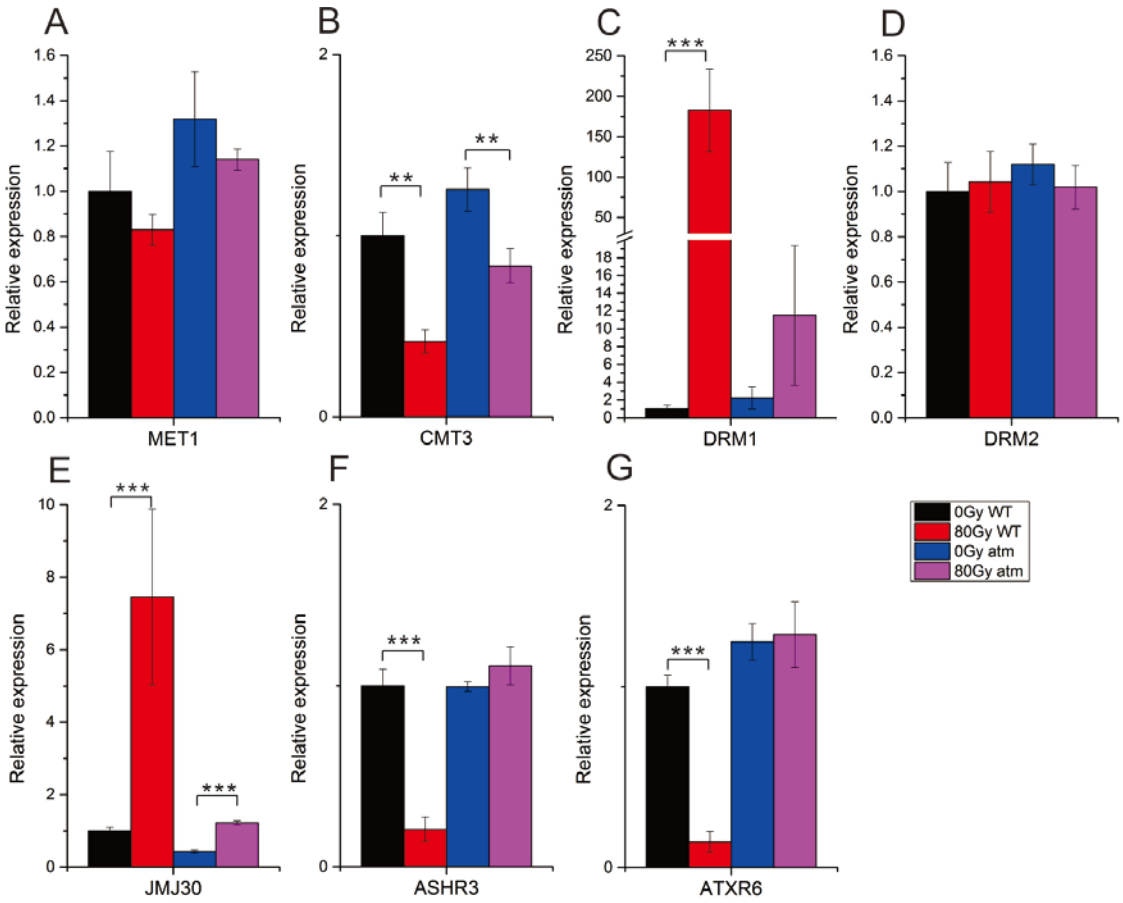


Figure 32: qRT-PCR detection of *MET1, CMT3, DRM1, DRM2, JMJ30, ASHR3*, and *ATXR6*. qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation. ** P < 0.01. *** P < 0.001 (t-test).

AT1TE69975 and AT1TE69990 inserted in *WRR4* introns, they co-expressed by the induction of X-ray induction in WT. It will be interesting to know whether DNA methylation is associated with their transcriptional regulation. In WT plants, AT1TE69975, AT1TE69990 and *WRR4* were all transcriptionally upregulated. However, bisulfite sequencing showed DNA methylation was different at these loci (Figure 33). In general, cytosine methylation level in AT1TE69990 was higher than in AT1TE69975 and the *WRR4* promoter region. By induction of X-ray, CG methylation was increased from 31 % to 39 %, but CHG methylation was reduced from 34 % to 25 % in AT1TE69975 three hours after irradiation. CHG methylation of the *WRR4* promoter was decreased by 10 % to 6 %, and there were no big changes of CG or CHH methylation after DSBs. Methylation of CG, CHG and CHH sequence contexts at AT1TE69990 showed various

reductions. CG methylation decreased from 100 % to 92 %, CHG methylation dropped 22 % from 44 %, and CHH methylation showed 15 % lost to 5 % in AT1TE69990. When taking a closer look at every cytosine site in analyzed fragments, it is more obvious that methylation levels in AT1TE69975 and promoter of *WRR4* were generally lower than in AT1TE69990 before irraidation (Figue 33). It also indicates that the transcriptional upregulation of AT1TE69990 was a result from the decreased methylation, especially non-CG methylation (Figue 33), after DSBs.

After investigating DNA methylation at specific loci, histone mofications were detected after DSBs. Western blot was executed with the same amount of proteins of each sample for analyzing different histone modification with anti-H3K4me2 (Millipore, 05-1338), anti-H3K4me3 (Abcam, ab9049), anti-H3K9me2 (Millipore, 05-1249), anti-H3K9ac (Abcam, ab10812), anti-H3K36me2 (Abcam, ab9049), and anti-Histone H3 (Abcam, ab1791) anti-serum. The results illustrate that there was no obvious alteration of detected histone modifications three hours after irradiation, neither in WT nor in the *atm* mutant plants (Figure 34). This indicates that in the early response to X-ray induced damage, there are no general histone modification changes in plants, at least the detected ones, despite the fact that there were few histone mofication genes showing transcriptional alterations.

3.2.9 AGOs in response to X-ray induced DNA damage

AGO4, an important factor of RdDM in plants, recruits small RNAs and further initiates DRM2-mediated *de novo* methylation at cytosine of non-CG sites (Zilberman et al. 2003). It controls transcriptional gene silencing of transposons (Creasey et al. 2014). Another interesting member of Argonaute proteins is AGO2. A recent report mentioned that AGO2 recruits DSB-induced small RNAs (diRNAs), which originate from DNA break-adjacent regions and mediate DSBR (Wei et al. 2012). A closer look at the members of this family could help to get a clue on the involvement of small RNAs in gene silencing after irradiation.

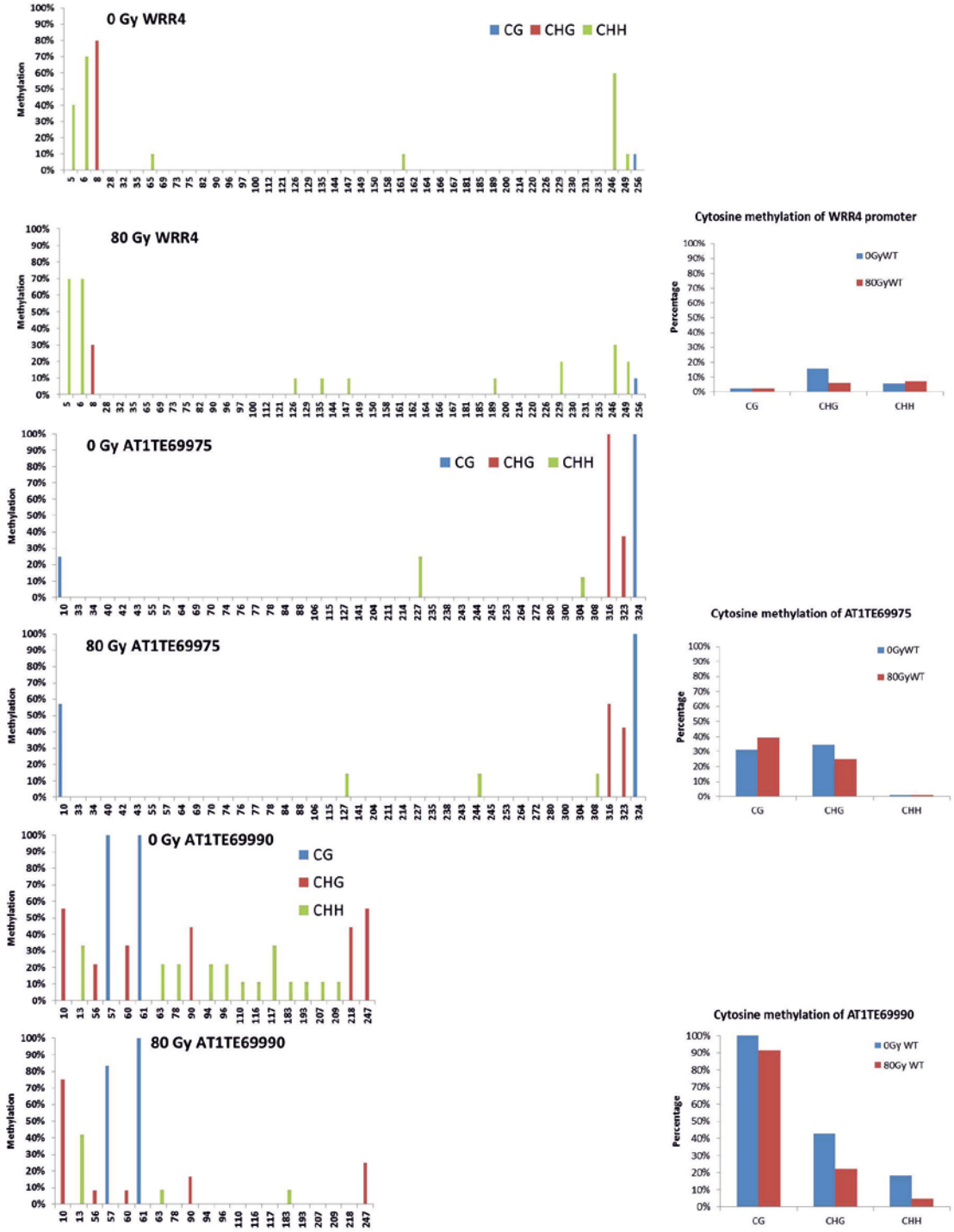


Figure 33: Cytosine methylation of AT1TE69975, AT1TE69990 and promoter region of *WRR4*. Methylation Percentage of each sample was calculated via inputting sequenced fragments on CyMATE (Cytosine Methylation Analysis Tool for Everyone).

On the left, bar charts with numbers show methylation level of each cytosine in analyzed sequences. Each number represents one cytosine postion. Blue, red and green colors point to CG, CHG, and CHH contexts. On the right, summerized methylation percentages of *WRR4* promoter region, AT1TE69975, and AT1TE69990 are shown in bar charts.

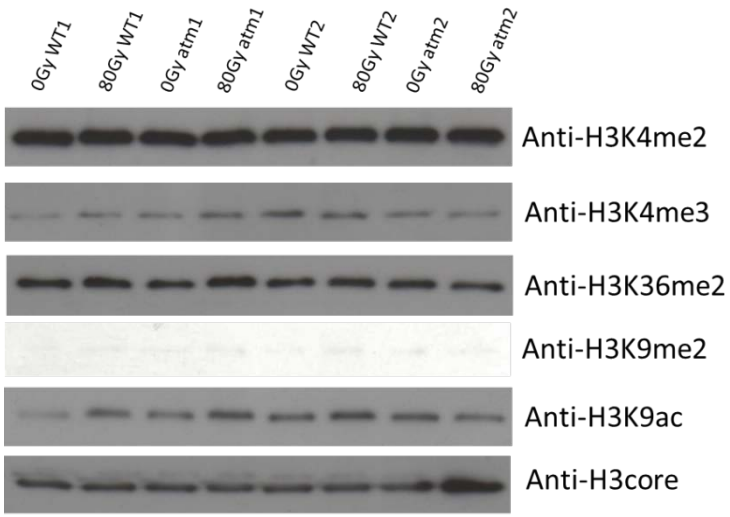


Figure 34: Histone H3 modifications in WT and *atm* mutant 3 hours post X-ray treatment.

Western blot was done with three biological replicates (Two are shown here). All first anti-sera were with 1:1000 dilutions and 1:2000 dilutions for second anti-serum. Histone 3 was used as loaded control here.

In RNA-seq data, *AGO4* expression remained stable after X-ray treatment. *AGO2* showed a 4-fold increase in WT three hours after irradiation. The second induced Argonaute family member by X-ray was *AGO7*, but it was only 1.5-fold upregulated in WT plants. No other Argonaute genes showed a significant transcriptional alteration. The expression of *AGO2*, *AGO4* and *AGO7* were also confirmed by qRT-PCR (Figure 35).

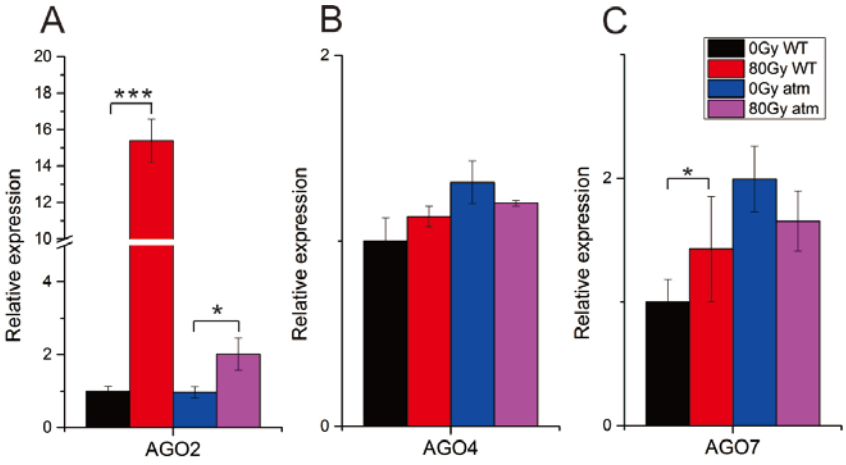


Figure 35: qRT-PCR detection of *AGO2*, *AGO4* and *AGO7*.

qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation. * P < 0.05. Three stars: ** P < 0.001 (T-test).

The unchanged transcriptional level of *AGO4* suggests that there was no dramatic alteration of RdDM induced *de novo* DNA methylation three hours after X-ray in plants. The upregulation of AGO2 fits its function in DSBR. *AGO7* participates in an antiviral RNA silencing pathway (Qu et al. 2008) and is required for trans-acting siRNAs (ta-siRNA) biogenesis (Jouannet et al. 2012). Its role in damage repair is still exlusive.

In summary, RNA-seq revealed more comprehensive transcriptomic changes three hours after X-ray induced DSBs in *Arabidopsis*. ATM not only regulated damage repair genes in response to irradiation, but also regulated immune genes, stimulus response genes, such as ethylene-responsive and stimulus of oxygen level, and even heat shock proteins. In a group of downregulated genes, many biological processes were categorized, which could improve the understanding of genes in response to irradiation induced DSBs. An interesting observation was that TEs/repeats were both activated and suppressed in the early response to X-ray induced damage, and this process was regulated by ATM as well, although overall, transcription of TEs and repeats was less frequently affected by DNA damage than that of normal genes. Even more intriguingly, RNA-seq revealed TEs/repeats derived lncRNA were in response to DSBs in an ATM-dependent manner.

3.3 Further analysis of DRM1, JMJ30 and AGOs in response to DSBs

3.3.1 Genotyping of selected mutants

In order to further investigate the roles of regulated genes involved in epigenetic processes and in DNA damage response (*DRM1*, *JMJ30* (*JMJD5*), *AGO2*, and *AGO7*), their T-DNA mutant lines were selected. *DRM1* homolog gene, *DRM2*, and another Argonaute protein, *AGO4*, were also chosen. As an extremely DNA damage-upregulated gene, *BRCA1* was chosen as well (gene structures and T-DNA inserted positions are shown in figure 36).

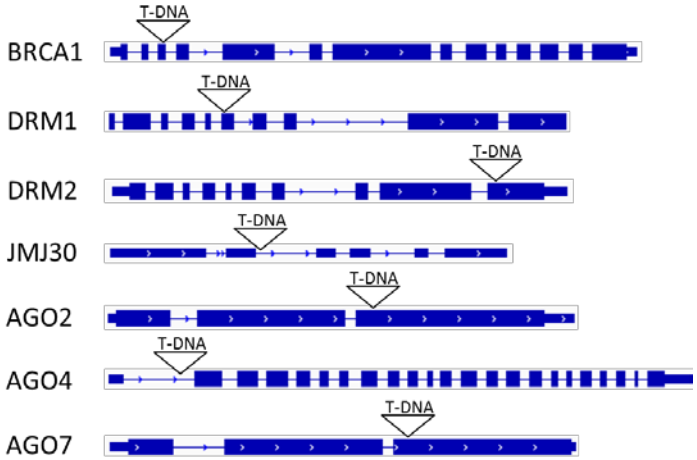


Figure 36: Gene structures and T-DNA insertions.

T-DNA inserted in the 3rd exon of *BRCA1*, 6th exon of *DRM1*, 10th exon of *DRM2*, 2nd intron of *JMJ30*, 3rd exon of *AGO2*, 1st intron of *AGO4*, and 3rd exon of *AGO7* of selected mutant lines, respectively. Blocks represent exons and lines mean introns.

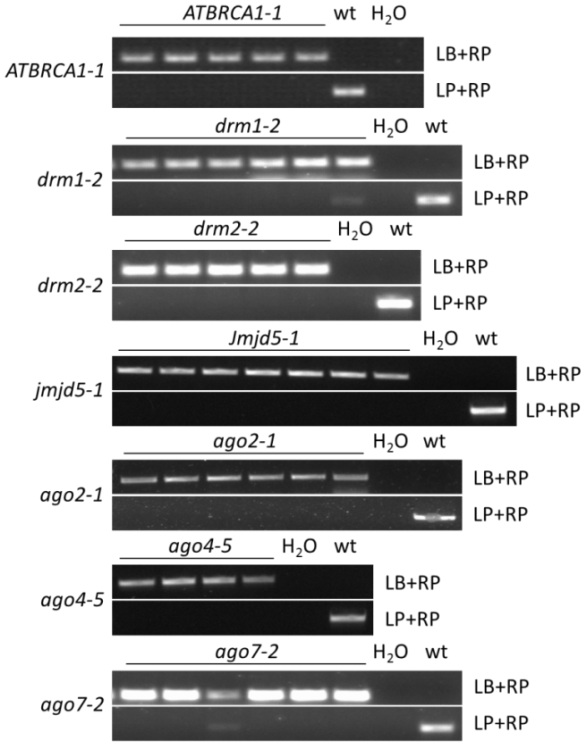


Figure 37: Genotyping PCR of T-DNA mutant lines.

Electrophoresis separated PCR products in 1 % agarose gel with ethidium bromide. The primer pair LB and RP amplified the T-DNA insertion, LP and RP amplified the intact gene fragment. The sample only presented a band amplified by using LB and RP was homozygous mutant plant. H_2O: water control; WT: WT control. Several independent plants of each mutant were detected. Plant showed only LB+RP product was homozygous mutant.

Homozygous T-DNA mutant lines *ATBRCA1-1* (SALK_014731), *drm1-2* (SALK_021316), *drm2-2* (SALK_150863), *jmjd5-1* (SAIL_811_H12), *ago2-1* (SALK_003380), *ago4-5* (WISCDSLOX338A06), and *ago7-2* (SALK_095997) were identified by genomic genotyping PCR. Two PCR reactions were performed for each T-DNA plant. LB and RP amplified T-DNA insertion fragment, LP and RP amplified intact genomic sequence of the specific gene (Figure 37). Primers used for the PCR are listed in table S4 (Appendix). Homozygous mutant plants were propagated and used for subsequent X-ray treatment.

3.3.2 Mutants responded differently to X-ray irradiation

Mutant plants were grown on half MS medium plates for 2 weeks, then treated with 80 Gy dosage X-ray radiations. Seedlings were transferred into soil and kept growing under long day conditions in greenhouse.

30 days after X-ray treatment, irradiate WT plants were smaller than unirradiated WT plants. They showed shorter shoots and small rosette leaves. The *atm* mutant plants were almost dead 30 days post ionizing treatment. The growth of *brca1* and *drm1* was retarded compared with WT plants, which indicated that they were more sensitive to irradiation than the WT. Surprisingly, *jmj30*, *ago2*, and *ago4* were much more tolerant to X-ray than the WT, and their irradiated and un-irradiated plants did not differ a lot. The residual two mutant lines, *drm2* and *ago7*, did not show obvious difference to WT after irradiation (Figure 38).

BRCA1 is known as a regulator of DNA repair. Previous research reported that DRM1 was a homolog of DRM2 with weak methyltransferase ability in *Arabidopsis* (Cao and Jacobsen 2002). However, the novel phenotype induced by X-ray, which is similar to *BRCA1* defective plants, suggests it might be involved in the DNA repair process as well.

The phenotypes of *jmj30*, *ago2*, and *ago4* generated by X-ray treatment are also interesting. JMJ30 is an *Arabidopsis* circadian regulator (Lu et al. 2011b). Its mouse

ortholog JMJD5 was shown to be able to demethylate di-methyl H3K36 (Ishimura et al. 2012). *AGO2* is a plant virus defending gene (Harvey et al. 2011). It also plays a role in DSBR (Wei et al. 2012). *AGO4* is an RdDM gene (Zilberman et al. 2003, Qi et al. 2006), and it suppresses small transposons by cooperation of chromatin and siRNA pathways (Tran et al. 2005). Their tolerance to X-ray might suggest they have yet uncharacterized functions in DNA damage response.

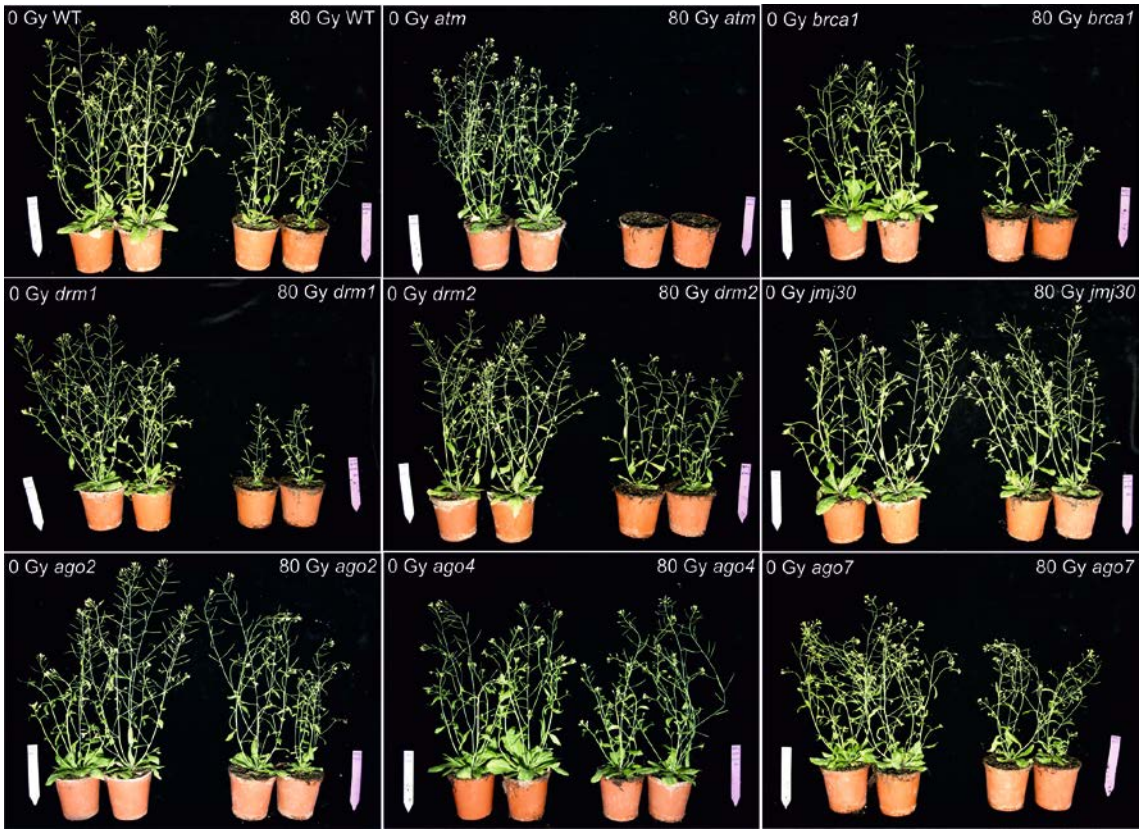


Figure 38: Phenotypes of mutant plants post irradiation.

Unirradiated and irradiated plants are on the left and right side, respectively. 30 day after X-ray treatment, *atm* showed severe sensitivity to damage; *brca1* and *drm1* mutants were more sensitive compare to WT; while *jmj30*, *ago2*, and *ago4* were surprisingly more tolerant than WT; *drm2* and *ago7* acted similar to WT.

3.3.3 DRM1 and BRCA1 affect TEs and lncRNAs in response to DSBs

BRCA1 is a DSBR gene which functions in HR (Yoshida and Miki 2004). It facilitates

resection of DNA ends at breaks and subsequently initiates HR to recruit other proteins in DSBR (Schlegel et al. 2006, Yun and Hiom 2009). Its mutant showed severe sensitivity to induced DSBs (Reidt et al. 2006). The similar phenotype of *DRM1* and *BRCA1* defective plants after X-ray suggests their potentially similar functions or different functions but interrupt the same pathway which results in a similar phenotype in response to breaks.

In parallel to observed phenotypes, total RNA was extracted from *drm1* and *brca1* mutant plants 3 hours post X-ray irradiation for analyzing transcripts induced by damage. Differentially expressed TEs and lncRNAs candidates were detected by qRT-PCR.

qRT-PCR showed that the upregulations of AT1TE04710 and AT1TE04720 were diminished both in *brca1* and *drm1* mutants compared with WT plants (Figure 39). LncRNA RLFS_026432 (annotated by PLncDB) was increased approximately 500 and 400-fold by the induction of X-ray in *brca1* and *drm1,* respectively. This is much higher than the 250-fold increase in WT. In contrast to RLFS_026432, At5NC029040 (Chr5:7891008-7891624) was downregulated by radiation, but its transcript kept a much higher level in *brca1* and *drm1* (Figure 39). The results of qRT-PCR indicate that *BRCA1* and *DRM1* both control these TEs and lncRNAs in response to X-ray. Combined with the retarded phenotypes of *brca1* and *drm1* post irradiation (Figure 38), the results suggest that *DRM1* is involved in the DDR process as well as *BRCA1*.

Interestingly, transcriptional levels of AT1TE69975, AT1TE69990, *CACTA-like* element (AT2G14230), *LINE* (AT2G11240), and At5NC029040 were higher in unirradiated *brca1* mutant than WT plants (Figure 39). This means *BRCA1* might have a role of suppressing TEs and lncRNAs.

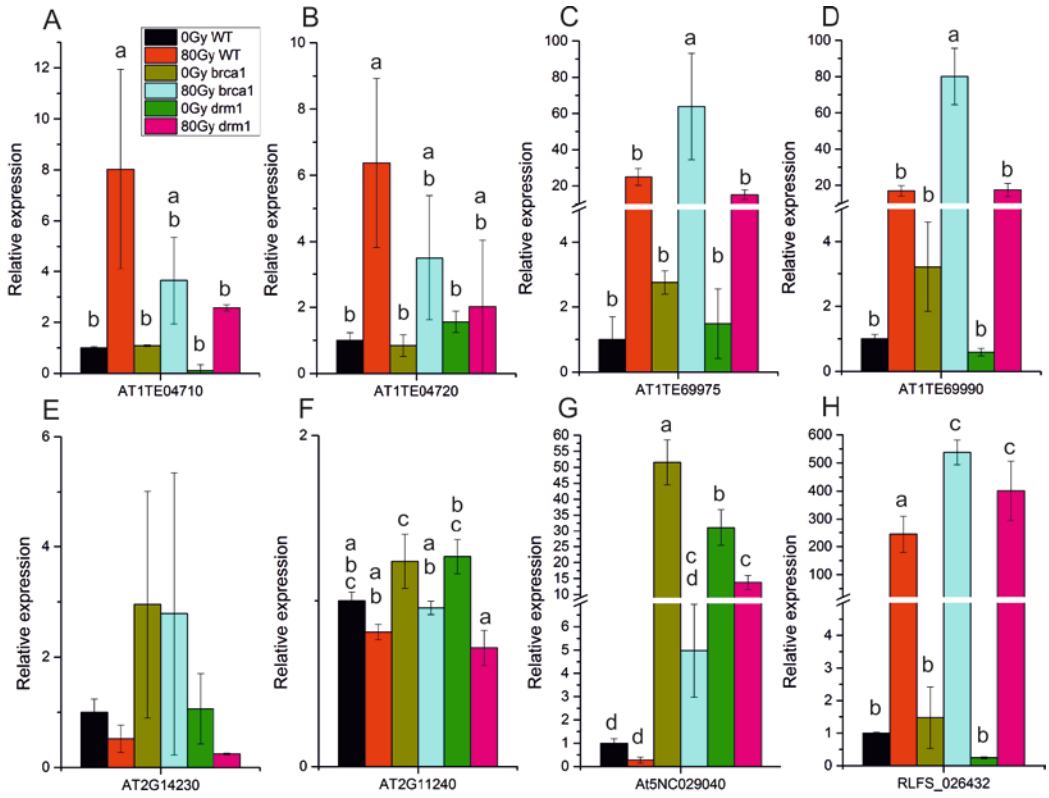


Figure 39: Transcripts of TEs and lncRNAs detected in *brca1* and *drm1* mutants by qRT-PCR.

A – D: AT1TE04710, AT1TE04720, AT1TE69975 and AT1TE69990 are TEs that are also predicted to be non-coding transcripts. E, F: AT2G14230 and AT2G11240 are TE genes. G, H: At5NC029040 and RLFS_026432 are lncRNAs annotated by PLncDB. qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation. Statistic analysis was performed always among different samples of one transcript. Relative expression levels with different letters were significantly different according to ANOVA followed by Tukey test.

3.3.4 Regulation of lncRNAs and TEs in *JMJ30*, *AGO2*, and *AGO4* defective plants

The epigenetic functions of JMJ30, AGO2, and AGO4 have been described before. After X-ray induced DSBs, *jmj30*, *ago2* and *ago4* mutants were unexpectedly tolerant to damage compared with WT (Figure 38). It would be interesting to know if they regulate TEs or lncRNAs in response to DSBs. The same TEs and lncRNAs mentioned in 3.3.3 were detected also in *jmj30*, *ago2* and *ago4* mutants. Although these mutants shared a similar resistance to X-ray, qRT-PCR showed that they regulated TEs and lncRNAs differently (Figure 40).

After X-ray irradiation, the *CACTA-like* element (AT2G14230) was particularly

upregulated in the *jmj30* mutant. *JMJ30* was specifically required for transcription of lncRNA RLFS_026432, but its absence did not affect the DSB-induced transcriptional activation of this lncRNA. AT1TE69975 showed a similar transcriptional level after irradiation in WT and *jmj30* mutants. However, its activity seems to be suppressed by *JMJ30* under normal conditions. AT1TE69990 is located in *WRR4* intron as well as AT1TE69975 (Figure 30). The qRT-PCR results suggested that it was not suppressed by *JMJ30* under normal conditions, whereas, interestingly, its damage induced transcriptional activation was decreased in *jmj30* mutant compared with WT.

AT1TE69975 was suppressed by *AGO2* under normal conditions as well. The other suppressed transposable element by *AGO2* is AT1TE04720. Its transcriptional level in unirradiated *ago2* was higher than in WT, *jmj30* and *ago4* mutant. However the activation of AT1TE04720 was not impacted by *AGO2*, as well as *JMJ30* to AT1TE69975. Interestingly, compared with WT, *AGO2*-defective plants showed a higher transcription of AT1TE69975 both before and after X-ray treatment. Lack of *AGO2* also led to a higher level of RLFS_026432 after induced damage (Figure 40).

The *AGO4*-defective plant showed the highest upregulation of AT1TE69975, AT1TE69990 and RLFS_026432 after irradiation (Figure 40). *JMJ30* and *AGO4* both inhibit the transcription of AtNC029040 in unirraidated plants. An X-ray irradiation insensitive TE, AT2G04460 (*COPIA95*), was also observed. It was highly upregulated in *ago4* mutant plants, but with no difference before and after irradiation. These observations indicate *AGO4* might play a role in inhibiting part of TEs or lncRNAs in response to DSBs.

Although *jmj30*, *ago2*, and *ago4* mutant plants had similar resistant phenotype after X-ray treatment, their various regulations of TEs and lncRNAs suggested a complex network in response to DSBs.

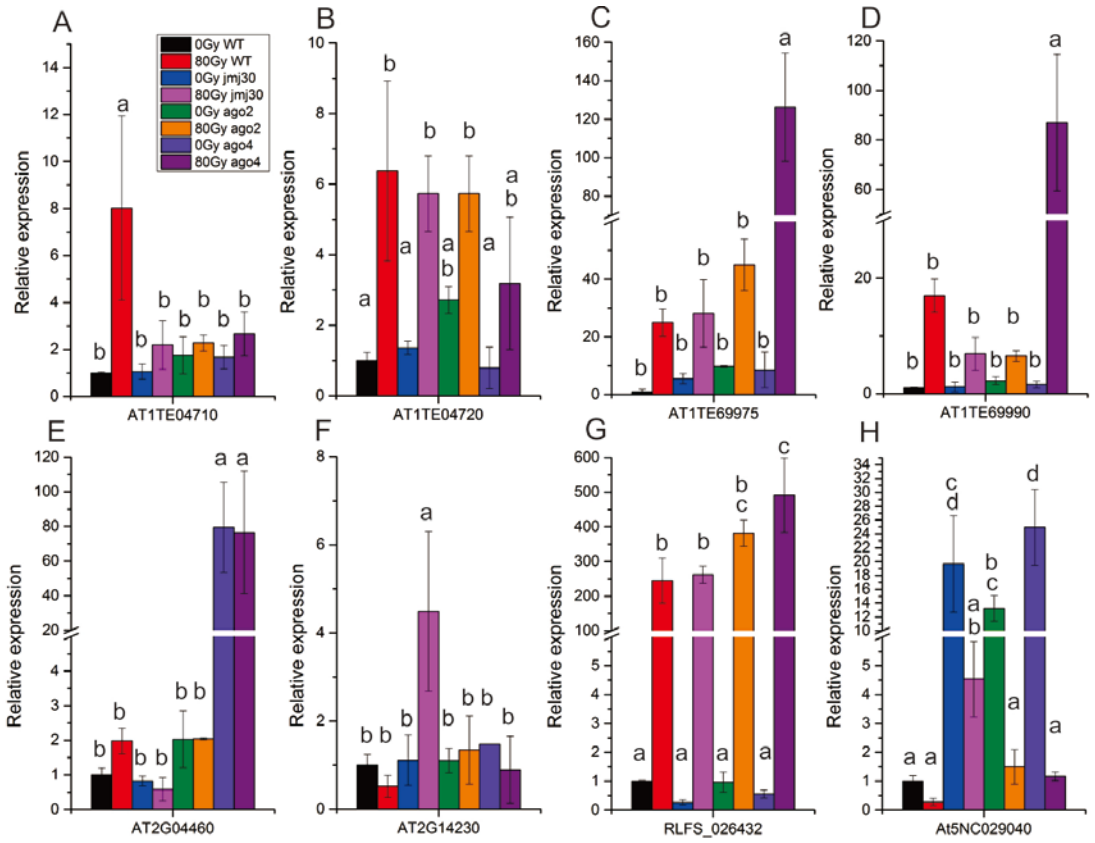


Figure 40: Detected transcriptional levels of TEs and lncRNAs in *jmj30*, *ago2*, and *ago4* mutants.

A – D: AT1TE04710, AT1TE04720, AT1TE69975 and AT1TE69990 are TEs that are also predicted to be non-coding transcripts. E, F: AT2G04460 and AT2G11240 are TE genes. G, H: RLFS_026432 and At5NC029040 are lncRNAs annotated by PLncDB. qRT-PCR was done with three biological replicates. Expression levels were normalized to beta-actin. Error bar: standard deviation. Statistic analysis was performed always among different samples of one transcript. Relative expression levels with different letters were significantly different according to ANOVA followed by Tukey test.

DRM1, *BRCA1*, *JMJ30*, *AGO2* and *AGO4* were investigated if these genes, or proteins, have a role in the overcoming of X-ray-induced DNA damage. *DRM1* and *BRCA1* defective plants show severe sensitivity to X-ray and some lncRNAs have similar regulation patterns in both mutants. This suggests that *DRM1* might play a novel role in DSBs response, and it probably participates in a similar pathway as *BRCA1* to control lncRNAs. Although *jmj30*, *ago2* and *ago4* mutants all showed more tolerance to irradiation, they regulate lncRNAs diversly. Despite a lack of confirmatory evidence, qRT-PCR results indicate a complex regulating network with *DRM1*, *BRCA1*, *JMJ30*, *AGO2* and *AGO4* participating in multiple pathways which regulate TEs/repeats derived

lncRNAs downstream of ATM in response to DSBs.

3.4 Regulation of heterologous TE

Endogenous TEs respond to environmental and cellular stresses. In the above work, a small fraction of TEs were either activated or suppressed by X-ray induced DSBs in plants, but compared with normal genes they seemed less sensitive to DNA damage. LncRNAs derived from TEs/TE-related fragments in response to damage suggests endogenous transposable elements/TE-related regions might play important roles in the anti-stresses process. Although epigenetic machinery and other factors have been known in regulating TEs/TE-related regions, however, many questions about the functions of lncRNAs originating from these loci still remain.

Besides endogenous TEs, it would be interesting to know how heterologous transposons are regulated in cells. TEs from endogenous or heterologous organisms have been used as genetic tools. *Ac/Ds* system is a widely used transposon model in plants. However, its germinal excision frequencies differ greatly in different target plants. *Ac* germinal transposition frequency in *Arabidopsis* is only 1 %, or even less. DNA methylation and negative dosage effect are possible explanations for the low ratio. So far, the negative dosage effect of *Ac/Ds* has not been observed in *Arabidopsis*, and how fast *Ac/Ds* are silenced via DNA methylation is still unclear. On the other hand, incorrect splicing and early terminated *Ac* transcripts were reported in *Arabidopsis* and sugar beet (Jarvis et al. 1997, Lisson et al. 2010). In this section of work, point mutations were created at reported locations with the aim of preventing mis-processed transcripts.

3.4.1 Generating *Ac* mutants and transgenic plants

According to the sites of cryptic introns and early terminated transcripts discovered in previous reports (Jarvis et al. 1997, Lisson et al. 2010), two constructs with 3 (*p35S::cAc-M3*) and 7 (*p35S::cAc-M7*) point mutations in *AcTPase* CDS without

changing the amino acid sequence were generated, respectively. These two vectors were aimed to prevent those formerly detected mis-processed transcripts of *Ac* in plants. Another construct containing WT *AcTPase* CDS (*p35S::cAc*) was created as a control (Figure 41). T-DNA binary plasmids were transformed into *Arabidopsis* via *Agrobacterium*. Positive transformed plants were first selected on half MS medium plants with Hygromycin B (15 mg/L). Transgenic plants were then confirmed by genome PCR and plants with a single insertion of *AcTPase* T-DNA were further identified by Southern blot (Figure 42). The probe used for Southern blot was generated by PCR via labelling DIG-11-dUTP to *Ac* fragment (position at *Ac* genomic DNA: 1239 - 1688).

3.4.2 The *miniDs* excision assay

Plants with single T-DNA insertions of WT *AcTPase* CDS (cAc), 3 mutations of *Ac* (M3), and 7 mutations of *Ac* (M7) were crossed with *miniDs* plants (3-44-2, generated by My-Linh Doll). Their progenies were selected on half MS medium plates within Hygromycin (15 mg/L) and Kanamycin (50 mg/L). 2-week old seedlings that survived on the selection plates were then transferred to soil and kept growing for another two weeks. Genomic DNA was isolated from individual plants of each of the hybrid lines. The excision events in the plants were investigated by PCR. Primers used for this PCR could amplify the fragment containing the whole *miniDs* element and its flanking regions. When excision occurred, a smaller product without the *miniDs* would be generated. The transposition frequency in each AcTPase X mDs line was calculated by the number of plants which showed a small product divided by the whole population of this line.

PCR results obtained from two WT AcTPase X mDs (cAc_6 and cAc_9), three M3 X mDs (M3_13, M3_14, and M3_15), and three M7 X mDs (M7_1, M7_2, and M7_3) plants illustrated that in WT AcTPase plants there were much more *Ds* excisions than in either M3 or M7 plants (Figure 43). The transposition frequencies in cAc_6 X mDs and cAc_9 X mDs plants were 19.6 % and 37.74 %, respectively. However, the highest ratio

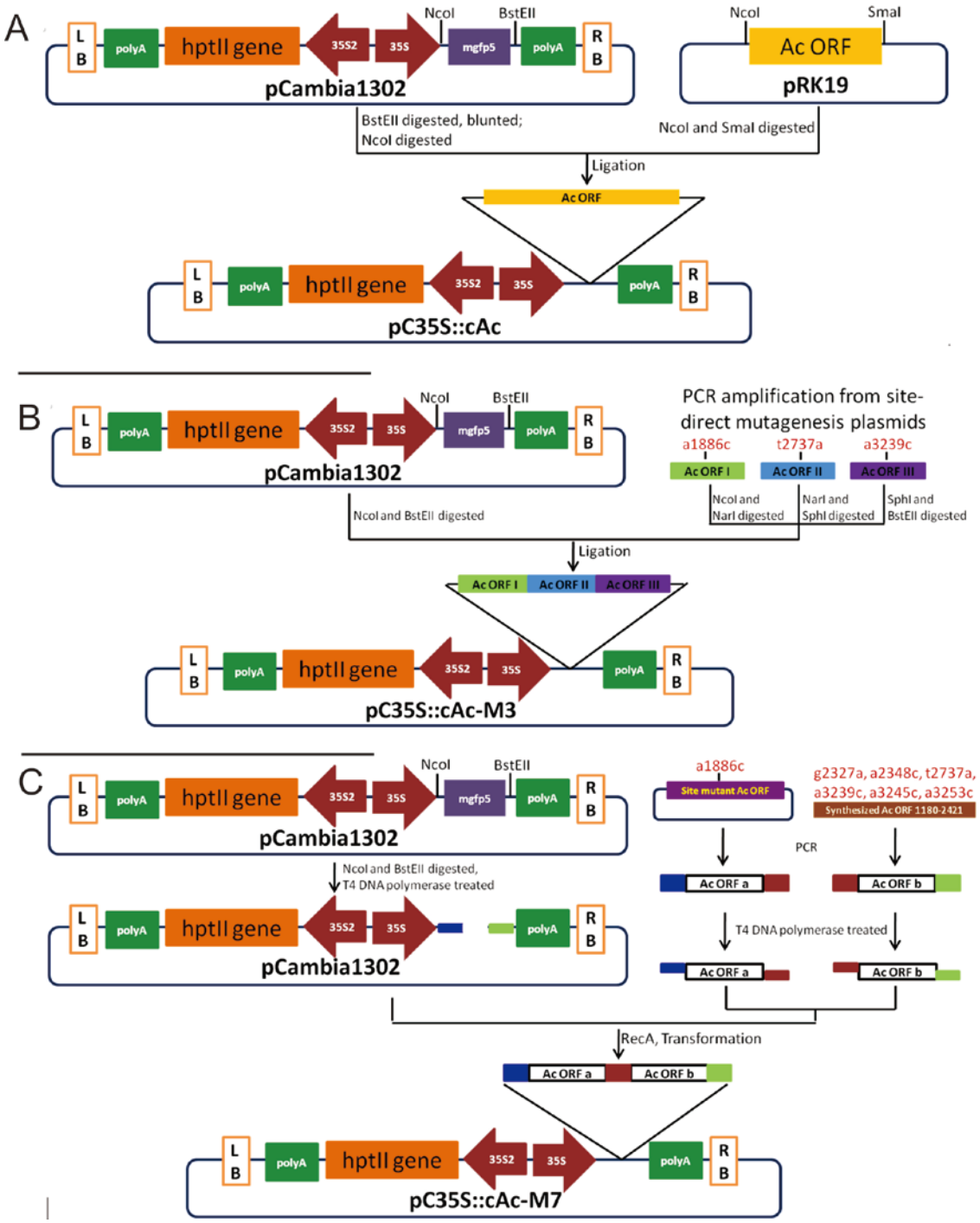


Figure 41: Construction of *Ac* cDNA derivatives with eliminated mis-processing sites.

(A) *pC35S::cAc* was generated by inserting the full length WT *Ac* transposase ORF in pCambia1302. Fragment of *Ac* transposase ORF was cleaved from pRK19 with NcoI and SmaI, then inserted between NcoI and BstEII sites in pCambia1302. (B) *pC35S::cAc-M3* was constructed by fusing three ORF segments that contain each 1 point mutation (a1886c, t2737a, and a3239c) created by site-direct mutagenesis. Fragments containing a1886c, t2737a, and a3239c were ligated in the order NcoI-SphI-BstEII, and then assembled to pCambia1302 between NcoI and BstEII sites. (C) *pC35S::cAc-M7* was assembled from two ORF fragments by sequence and ligation independent cloning (SLIC). *Ac* ORF fragment 'a' has 1 mutation (a1886c) created by site-directed mutagenesis. *Ac* ORF fragment 'b' contains 6 point mutations (g2327a,

a2348c, t2737a, a3239c, a3245c, and a3253c) and was generated by *de novo* synthesis. These two fragments were ligated and inserted between NcoI and BstEII sites in pCambia1302 via homologous recombination.

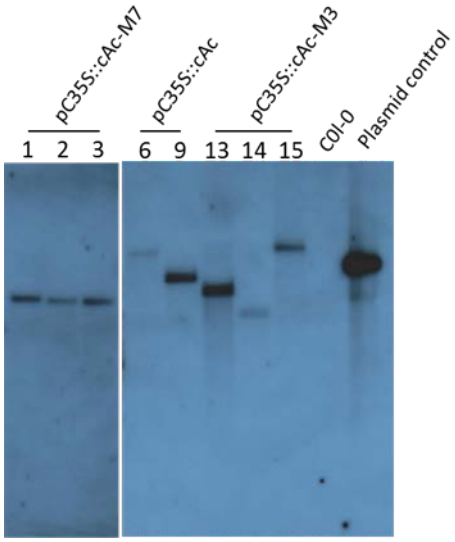


Figure 42: Southern blots of *AcTPase* WT and derivatives plants.

Transgenic plants with single T-DNA insertions of WT *AcTPase* CDS (pC35S::cAc), 3 mutations of Ac (pC35S::cAc-M3), and 7 mutations of Ac (pC35S::cAc-M7) were identified by Southern blot. 2 µg genomic DNA from each sample was digested with BamHI overnight. The amount of plasmid control (pC35S::cAc) loaded to the gel was 150 pg.

in mutant plants was 14.29 % in M3_15 X mDs and even decreased to 0 % in M3_13 X mDs, M7_2 X mDs, and M7_3 X mDs plants (Table 15).

Semi RT-PCR revealed that those mis-transcripts were not observed in mutant plants (Figure 44, lanes 5 – 10). In M3_14 X mDs_44, M3_14 X mDs_45, and M7_1 X mDs_2 mutant plants there were less amount of *Ac* transcripts than in other detected plants (Figure 44). However, the amount of *Ac* transcript did not have a positive relationship with excision. For instance, excision happened in both cAc_6 X mDs_11 and 5 M3_14 X mDs_44, although the amount of *Ac* transcript showed a large difference between them. No excision existed in cAc_6 X mDs_10, but it possessed a high *Ac* transcriptional level (Figure 44).

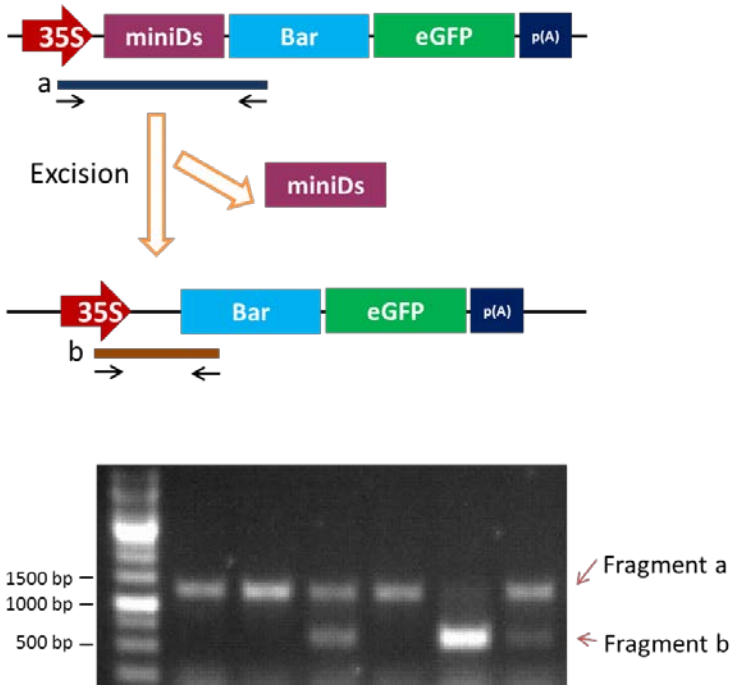


Figure 43: *Ds* excision screening.

Top: Schematic diagram of *Ds* excision. *MiniDs* element is inserted between 35S promoter and Bar gene. Primer pair 35S and bar2 (pair of black arrows) is used for amplifying the region including *miniDs*. When *miniDs* does not excise from its original position, a 1279 bp fragment (fragment a, blue frame) will be generated by PCR. If *miniDs* excised from its original location, an approximately 650 bp product (fragment b, dark orange frame) will be amplified by PCR. Down: an example of PCR result of *AcTPase* X *miniDs* plants. The lanes showed fragment b were counted as *Ds* excised.

Table 15: *Ds* transposition efficiencies in transgenic plants

Samples	**Total number**	**Number of excision**	**Transposition efficiencies**
cAc_6 X mDs	51	10	19.61 %
cAc_9 X mDs	53	20	37.74 %
M3_13 X mDs	53	0	0.00 %
M3_14 X mDs	44	2	4.55 %
M3_15 X mDs	49	7	14.29 %
M7_1 X mDs	39	2	5.13 %
M7_2 X mDs	54	0	0.00 %
M7_3 X mDs	53	0	0.00 %

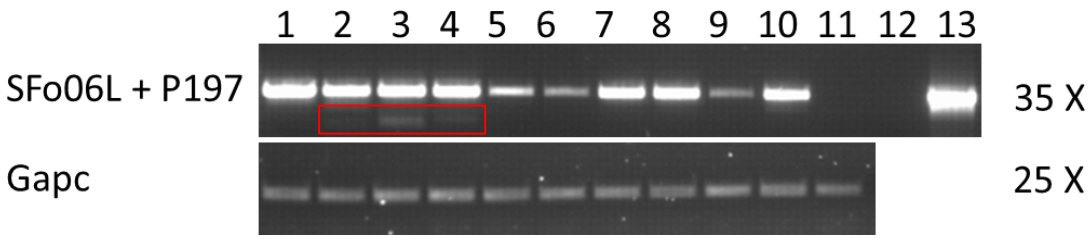


Figure 44: Detected *Ac* transcripts in selected plants by semi RT-PCR.

1 cAc_6 X mDs_10 (N), 2 cAc_6 X mDs_11 (E), 3 cAc_9 X mDs_3 (E), 4 cAc_9 X mDs_4 (N), 5 M3_14 X mDs_44 (E), 6 M3_14 X mDs_45 (N), 7 M3_15 X mDs_47 (E), 8 M3_15 X mDs_50 (N), 9 M7_1 X mDs_2 (N), 10 M7_1 X mDs_8 (N), 11 WT, 12 ddH_2O, 13 plasmid control. N represents no excision, E means excision. SFo06L and P197 could amplify a 1912 bp fragment from Ac transcript which covers all the point mutations. *Gapc* was used as reference gene. Bands in red frame: alternative transcripts detected in WT *Ac* plants.

In this part of the work, transgenic plants transformed with point mutated *AcTPase* CDSs, in order to prevent mis-processed *Ac* transcripts, were generated. Despite the mutations reduced the occurence of mis-processed *Ac* transcripts (Figure 44, lanes 5 – 10), *Ds* excision frequency was not improved in mutant plants. It suggests that incorrect splicing of *Ac* transcripts might not be the only regulation leading to low germinal excision of heterologous *Ac/Ds* elements in *Arabidopsis*.

4 Discussion

4.1 Meganuclease induced DSBs

4.1.1 Estradiol induced expression of meganuclease could control the generation of DSBs in *Arabidopsis*

Meganucleases are a class of endonucleases with very long DNA recognition sequences (12 - 45 bp) (Thierry and Dujon 1992). Homing endonucleases are naturally occurring meganucleases encoded by the introns and inteins of bacteria, phages, and some eukaryotes, such as *S. cerevisiae* and *Chlamydomonas reinhardtii* (Belfort and Roberts 1997, Jurica and Stoddard 1999, Chevalier and Stoddard 2001). I-Sce I and I-Cre I are two homing endonucleases derived from *S. cerevisiae* and *C. reinhardtii*, respectively (Dujon 1980, Rochaix et al. 1985).

Besides the natural meganucleases, Zinc-finger nucleases (ZFNs) are artificial restriction enzymes that are created by fusing the DNA-cleavage domain, typically the FokI domain, and zinc finger DNA-binding domains (Kim et al. 1996). The well-defined Zinc-finger nuclease QQR can cleave a short sequence (6 bp) between two GGGGAAGAA (5'-3') binding sites (Bibikova et al. 2001, Lloyd et al. 2005). Another custom made ZFN, ZFN3, can recognize the 9-bp long sequence GGTGGTGCT (5'-3') (Tovkach et al. 2009).

Because of their long target sites, both meganucleases and artificially generated ZFNs are rare cutting enzymes in organisms. They were used for creating DSBs in a target organism and investigating further the responses to DNA damage or being applied for gene targeting. For instance, I-Sce I has been used successfully for DNA repair research in mammalian cells and plants (Sargent et al. 1997, Puchta 1999b, Puchta 1999a, Wei et al. 2012). Another meganuclease, I-Cre I, has been adopted as a useful tool for transgene insertion and gene therapy (Arnould et al. 2011).

In this work, two *I-Sce I* CDSs, yeast original and *Arabidopsis* optimized (*ISI-opA*), and

two ZFNs (*QQR* and *ZFN3*) gene sequences were transferred into *Arabidopsis* plants under the control of an estradiol inducible promoter. The promoter is made up of eight copies of the *LexA* operator and the -46 35S minimal promoter. Its downstream genes will be transcribed only in the presence of estradiol-activated XVE (Zuo et al. 2000).

The strictly estrogen dependent expressions of *I-Sce I, ISI-opA*, *QQR*, and *ZFN3* were confirmed by qRT-PCR. It means that the estradiol inducible element can temporally modulate the transcription of all rare cutting nucleases. However, only the *ISI-opA* encoded I-Sce I enzyme generated obvious DSBs in transgenic plants. There was no previous report comparing the damage creating frequencies of these four sequences encoded nulceases in *Arabidopsis*, but the GUS staining and *in vitro* digestion assay indicated that *ISI-opA* encoded I-Sce I was the most efficient enzyme among these four to induce DSBs. The sequencing results also revealed that the dominant repair pattern of I-Sce I induced damage is deletion in *Arabidopsis*.

4.1.2 Meganuclease-induced DSB frequency is insufficient for detecting alterations in transposon activity or epigenetic status

Hitherto, I-Sce I has been used in plenty of previous researches. However, the efficiencies of damage varied in different studies. By using I-Sce I, 2 - 4 % of the cells generated a detectable GFP signal by HR in HeLa cells (Cuozzo et al. 2007), and up to a 10 % recombination ratio was observed in HEK-293 cells (Szczepek et al. 2007). In *Arabidopsis*, DSB and repair was observed in 5 - 10 % of cells (Prof. Dr. Holger Puchta, Universität Karlsruhe, personal communication). In this work, the detectable repair events was 6.9 % in estradiol induced ISI-opA and target site transgenic plants. The number of repair events which occurred in this work was reasonable when compared to those in former work, but it seems that the amount of damage in transgenic plants could not achieve the goal of inducing the activity changes of TEs when the qRT-PCR of TEs and genes candidates were considered.

As in other species, *Ku70* and *Ku80* are essential players of NHEJ in *Arabidopsis* as

well (Tamura et al. 2002). *LIG4* is another well-known gene that participates in the NHEJ process (van Attikum et al. 2003). *BRCA1* is important for keeping the genome stable, and it has a crucial function in HR (Shrivastav et al. 2008, Huen et al. 2010). RAD50 and MRE11 are components of the MRN complex (MRE11/RAD50/NBS1), which recognizes DSBs and recruits the important damage repair initiative kinase, ATM, to broken sites and leads to phosphorylation of the downstream repairing factors (Lee and Paull 2005). In this experimental setting, the NHEJ should be the main repairing process for DSBs, because the target site of the endonuclease was inserted right after the start codon of the *uidA* gene, and there was no homologous sequence for the interrupted fragment in plants. Though the DNA repair was observed, none of *Ku70*, *Ku80, and LIG4* was strongly regulated. The HR gene *BRCA1* and two of the MRN components, *MRE11* and *RAD50*, showed no obvious change. This result suggests that the background expressions of DDR genes might be enough to deal with the few amount of lesions created by the I-Sce I enzyme during the long induction time (48 hours).

In *Arabidopsis*, *ATMu1*, *ATCOPIA4*, *ATLINE1-4*, *AtLANTYS*, *ATGP1, ATENSPM2, HARBINGER, ATCOPIA72,* AT4G09480*,* and AT2G04460 were controlled by DNA methylation and histone acetylation. Some of them were regulated by H3K9 methylation and RNA directed DNA methylation (Lippman et al. 2003, Liu et al. 2011). *DDM1* regulates three *COPIA* elements (Tsukahara et al. 2009) and *MET1* and *CMT3* controlled *CACTA* family members were reported as well (Kato et al. 2003). In rice, DNA methylation and H3K9 di-methylation manipulated *Tos17* (Ding et al. 2007, La et al. 2011), and LINE1 was controlled by H3K4 tri-methylation (Cui et al. 2013).

In case those TEs were regulated after DNA damage, the DNA methylation and histone modification relative gene could also be altered transcriptionally. *MET1*, *CMT3*, *DRM2*, *DDM1* and *ROS1* were selected genes associated with DNA methylation. Some of histone methylation (*ATX1*, *ATX2*, *ATXR3*, *ATXR5*, *ASHR3*, *ASHH1*, *KYP*, *SUVH5*, *SUVH6*, and *SUVR4*) and demethylation (*JMJ14*, *JMJ15*, *JMJ30*, *IBM1*, *REF6*, and *EFS*) genes were chosen as detecting candidates. Histone deacetylase HDA6 was investigated as well. Unfortunately, none of them showed obvious regulation after induced damage in *Arabidopsis*. Seven TEs (*ATMUDR1*, *ATMUDR2*, *ATCOPIA4*,

gypsy-like element, *Mutator-like* element, *Spm (CAC1)*, and *ATLINE1-4*) were investigated by qRT-PCR in *ISI-opA* induced damage plants, four of them were slightly regulated and three of them were undetectable.

Because of the failure to detect changes of TEs activities and genes after induced DNA damage in transgenic plants, although the DNA cleaving system worked, this experimental system is not useful for investigating the relationship between DNA damage and regulation of TEs. The reasons for giving up this inducible system of further applications were: 1) the low damage ratio in whole plants could not induce remarkable transcriptional changes, which means the regulation might exist but is undetectable by qRT-PCR; 2) 48 hours induction of estradiol was a compromise for generating enough endonuclease, and caused a certain amount of damage, but it was not the best for observing a rapid response of TEs and genes to DNA breaks; 3) *ISI-opA* and target sites were transformed only in WT background plants. It is possible to observe more differentially regulated TEs or genes when compare WT with mutant lacking DDR gene after induction of DSBs.

4.2 Early responses to X-ray induced DNA damage in *Arabidopsis*

In order to obtain a large amount of DSBs and detectable rapid responses, X-ray irradiation device at Julius Kühn-Institut (JKI) in Quedlinburg was chosen as a damaging source. Besides WT plants, ATM defective plant were treated with X-ray as well. ATM is the essential kinase for initiating downstream protein to repair damage (Abraham 2003, Garcia et al. 2003). A new and powerful method, RNA-seq, was used to analyze the whole transcriptomic changes in early stage of DNA damage response. By comparing the next-generation sequencing data of WT and *atm* mutant plants, there would be more possibilities for novel discoveries of induced break responses in *Arabidopsis*.

4.2.1 RNA-seq reveals a complete overview of early transcriptomic responses to DSBs

Ionizing radiation (IR) induced damage has previously been investigated in *Arabidopsis* with various strategies. Some studies focussed on specific genes induced by IR. Gamma-irradiation induces expression of DNA ligase IV (LIG4), a protein that interacts with XRCC4 via BRCA C-terminal domains (West et al. 2000). Transcripts of POLY (ADP-RIBOSE) POLYMERASE genes (*PARP1* and *PARP2*) accumulate rapidly by induction of gamma-irradiation in *Arabidopsis*, which is in contrast to the post-transcriptional activation of the PARP-1 protein induced by DNA damaging reagents in mammalian cells (Doucet-Chabeaud et al. 2001). The understanding of IR induced DNA damage response was extended by the application of microarrays. After IR-treating Ler WT and ttg-1 (WD-40) mutant (*Ler* ecotype), signal transduction and transcription factors genes, such as *WRKY* and *MYB*, were activated mainly. Metabolism, cell rescue, defense, cell death and aging genes were also upregulated (Nagata et al. 2005). 5'-cis-element regions of the IR-response genes were identified in this work as well. By analyzing microarray data, hundreds of upregulated genes were discovered in response to IR, which are dependent on ATM but not ATR (Culligan et al. 2006). The above-mentioned strategies worked perfectly for identifying well-annotated genes, but are not suitable for discovering repetitive elements, such as TEs, or previously not annotated genes or transcripts.

In this work, total RNAs isolated from 80 Gy X-ray irradiated and unirradiated WT and *atm* mutant plants were used for RNA-seq. By using RNA-seq the same regulated genes were detected as using microarrays, but RNA-seq is more sensitive. The RNA-seq revealed additional ATM-dependent X-ray-inducible genes and processes, such as the AAA-type ATPase family proteins, calcium signaling, receptor kinase signaling, nitrilases, UDP Glycosyltransferases, Glutathione-S-transferases genes. Interestingly, auxin-responsive genes, especially the IAA family members, were strongly downregulated in WT but remained stable in *atm*, which by contrast, is opposite to

ethylene-responsive genes. In previous work, few differentially expressed WRKY factors were detected (Nagata et al. 2005, Culligan et al. 2006). Interestingly, in this study, RNA-seq detected 13 differentially expressed *WRKY* genes, and they were split into two groups in response to irradiation, upregulated in the WT but downregulated in mutant plants. It is worth to look for more detailed functions of those *WRKY* genes. After irradiation, the WT and *atm* mutant plants initiated different immune responses, the observation of WRKY factors acting in two groups indicate their particular functions in various immune pathways.

In addition to genes, RNA-seq detected TEs/repetas and novel transcripts not represented on microarrays. RNA-seq detected 116 TEs/repeats elements transcriptionally activated or suppressed after X-ray treatment. 179 novel transcripts were identified according to RNA-seq data. Of these, 9 and 8 were differentially expressed in WT and *atm* mutant after DNA damage and all these regulated novel transcripts were predicted to be lncRNAs.

Additional genes, TEs/repeats derived transcripts and novel transcripts detected in this work, indicate that RNA-seq is able to generate a more complete transcriptomic profile of DSBs response compared with previous microarray analysis.

4.2.2 TEs are insensitive to DSBs

In the past, several investigators have reported transcriptional activation and transpositional (re)activation of class I and class II TEs upon stress. Tobacco retrotransposon *Tto1* can be transcriptionally activated by tissue culture, infection, wounding, and defensive signal jasmonate. Its activation is controlled by a 13 bp cis-regulatory element in the promoter (Takeda et al. 1998, Takeda et al. 1999). Another tobacco retrotransposon *TntlA*, *copia-like* element, is activated by a pathogen (Grandbastien 2004). In *Antirrhinum majus* and *Arabidopsis*, the elevated activity of *Tam3* and copia78 is mediated by altered temperature (Hashida et al. 2003, Pecinka et al. 2010). Radiation and other DNA damaging agents can induce the activation of *Tn7* in *E.coli* (Shi et al. 2008), *Ty1* in yeast (Sacerdot et al. 2005), *mudrA* and *mudrB* in maize

(Questa et al. 2010), and *LINE-1s* in human cells (Farkash et al. 2006).

At first glance, all these examples seem to strengthen the impression that previously silenced TEs are transcriptionally upregulated and transpositionally reactivated when they suffer specific stresses by default. However, RNA-seq analysis reveals that the regulation of TEs induced by X-ray is bi-directional. Regulation of TEs is observed, however, when compared to the high number of differentially expressed genes, the fraction of transcriptionally up- or downregulated TEs is much lower. This implies that at least in young *Arabidopsis* seedlings TEs are rather less sensitive to stress by DNA damage than the normal genes.

This phenomenon leads to several hypotheses: (1) TEs are generally less sensitive to DSBs in early developmental stages, but are more prone to become activated later in development; (2) only a small number of TEs is responsive to DNA damage during early plant development; (3) both up- and downregulation by one stress could indicate that some TEs have more sophisticated functions during the response, and may not only act as passively manipulated targets.

In human glioma cells, proapoptosis genes become activated at 6 hours post-irradiation, whereas antiapoptosis genes are activated quite late, most of them later than 48 hours post-irradiation (Ma et al. 2013a). In *Arabidopsis*, transcripts of DNA repair genes *BRCA1*, *RAD51*, *RAD17* accumulated rapidly post irradiation. However, *LIG4* reaches the expression peak later than others at 6 hours post-irradiation (West et al. 2000, Garcia et al. 2003, Culligan et al. 2006). Short heat stress (SHS, 3 hours at 37 °C) causes upregulation of 8 TEs. 14 more TEs are induced by long heat stress (LHS, 30 h at 37 °C) in *Arabidopsis* (Pecinka et al. 2010). It seems that many genes and also TEs show a late response pattern. Therefore it is possible that more TEs become transcriptionally induced if later time points post irradiation were to be investigated. But considering the total number of LHS induced TEs, it is still a tiny proportion of the TE-related genome fraction in *Arabidopsis*, and so it is more likely that TEs are generally less sensitive to DSBs compared to genes.

4.2.3 LncRNAs are associated with DSBs response

In this study, ATM-dependent regulation of lncRNAs in response to X-ray induced DSBs in plants was shown for the first time. More interesting is that most of these lncRNAs are originated from TEs/repeats loci. Although the evidence of their specific roles in DNA damage is still absent, previous research probably can show some hints for following studies.

In plants, lncRNAs are supposed to be generated via Pol IV and Pol V machinery which share part of Pol II subunits when the biosynthesis is initiated (Wierzbicki et al. 2008, Ream et al. 2009). LncRNAs are speculated to (i) regulate post-translational modification of histones by acting as a *cis* or *trans* element to recognize the target locus, and recruit histone modification proteins, and participate in constructing a Pol II inaccessible compartment which leads to gene silencing; and to (ii) regulate transcriptional gene silencing through its dual function of creating siRNA as a precursor and load siRNA/AGO4 complexes to specific chromatin targets as a scaffold (Kato et al. 2005, Wierzbicki et al. 2009).

Xist and the locus at *FLC* are typical models of lncRNA functions in mammals and plants, respectively (Liu et al. 2007, Zhao et al. 2008, Liu et al. 2010b). The lncRNAs HOTAIR in human and COLDAIR in *Arabidopsis* suggested lncRNAs are associated with stress induced regulation (De Lucia and Dean 2011). Recently, Wan and colleagues detected that lncRNAs were responsive to DNA damage and were ATM dependent in mouse cells (Wan et al. 2013). This coincides with the observation in *Arabidopsis* in this study. The author named the most significantly induced lncRNAs as lncRNA-JADE because it is close to the *Jade1* gene in both human and mouse. LncRNA-JADE promotes the transcription of *Jade1* by interaction with BRCA1, and leads to increased histone H4 acetylation (Wan et al. 2013). LncRNA-JADE is highly conserved in mouse and human. However, no similar sequence is detected in *Arabidopsis*, but interestingly the *Arabidopsis* homolog AT3G14740 of mouse *Jade1* was downregulated 4-fold in WT three hours after irradiation.

The observation in this study indicates that ATM dependent regulation of lncRNAs in

response to DNA damage is conserved in both plants and mammals. However, the specific functions of lncRNAs might differ a lot between plants and mammals.

4.2.4 TEs/repeats-derived lncRNAs affect associated and nearby genes

In this work, 73 out of 91 and 8 out of 18 lncRNAs in WT and *atm* mutant originated from TEs/repeats. The locations of these transcripts on the genome distribute near the 5' end of gene coding regions, within the gene structure (exons and introns), close to gene 3' region, and also far away from genes in intergenic regions. 40 % and 28 % of regulated lncRNAs in wild type and atm mutants were co-expressed with their associated and near-by genes. For instance, AT1TE04710 and AT1TE04720 insert in front of AT1G05030. AT1TE69975 and AT1TE69990 insert in the first and fourth introns of *WRR4*, a disease resistance gene, respectively.

TEs/repeats can affect the expression of a nearby gene by altering the epigenetic status of that gene, interrupting the promoter, and read-through antisense transcription (Slotkin and Martienssen 2007, Wang et al. 2013). TE-derived fragments were observed in 25 % of human promoter sequences (Jordan et al. 2003). Domesticated LTR elements are found in promoters of neuronal apoptosis inhibitory protein (NAIP) family genes and they regulate the transcription of those genes in human, mouse, and rat (Romanish et al. 2007). Mice coat-color gene agouti is regulated by the epigenetic status of a proximal IAP retrotransposon, which generates an outward-reading transcript including the agouti gene (Blewitt et al. 2006). Another interesting color-controlling example is also subject to the regulation of methylation at the *MULE1* element which inserts into the promoter of DFR-B gene in Japanese morning glory (Iida et al. 2004). The above examples described TEs/repeats that regulate gene functions in a rather straightforward way, such as to interrupt promoter regions or alter DNA methylation status. It will be interesting to know if TEs/repeats derived lncRNAs affect gene functions on a new regulating layer.

A recent work showed 50 % of lncRNAs and coding RNAs expressed in the same direction in human bladder urothelial cell (Wang et al. 2014). In *Arabidopsis*, a fresh

paper showed that auxin induced gene *PID*, a key regulator of polar auxin transport, was regulated by its neighboring lincRNA APOLO (Ariel et al. 2014). It suggests that one possible function of lncRNAs is regulating associated or nearby genes in response to stress.

In subsequent work, it will be helpful to use knockout lines of lncRNAs to further investigate their functions in DNA damage response.

4.2.5 *DRM1* might be involved in DSBR and regulates lncRNAs downstream of *BRCA1*

DRM1 and *DRM2* are *Arabidopsis* homologs of mammalian *DNMT3s*, which *de novo* methylate cytosine. *DRM2* is expressed more strongly than *DRM1* in the WT *Arabidopsis* background (Cao et al. 2000). *FWA* (*FLOWERING WAGENINGEN*) transformation assay indicates that *DRM2* is the dominant *de novo* methylase, but not *DRM1* in *Arabidopsis*. The double mutant of *drm1drm2* is defective in *de novo* methylation of *SUP* (*SUPERMAN*) (Cao and Jacobsen 2002). However, this does not affect the maintenance of silencing *SUP* and *FWA* genes (Cao and Jacobsen 2002). Moreover, *DRM2* controlled *de novo* methylation is regulated by small RNAs (Cao et al. 2003, Pontes et al. 2006).

In the early response to X-ray, *DRM2* showed no obvious change, neither in *atm* mutant nor in WT plants. However, *DRM1* was significantly upregulated in irradiated WT and the upregulation in *atm* mutant was reduced. Interestingly, *drm1* mutant was sensitive to X-ray. Its retarded growth was similar to the irradiated *BRCA1* defective plants, whereas, *drm2* mutant responded to X-ray just as WT. qRT-PCR showed that *drm1* and *brca1* mutant had similar regulations of some selected TEs, AT1TE04710 and AT1TE04720, and lncRNAs, RLFS_026432 and AT5NC029040.

BRCA1 is a well-studied tumor suppressor gene. It regulates DNA damage induced activation of a cell cycle checkpoint and also participates in DDR, both NHEJ and HR (Wu et al. 2010). In mammalian cells, *BRCA1* regulates expression of *DNMT1* via binding to an OCT1 site (AACGTAA) in the *DNMT1* promoter, and further affects DNA

methylation globally (Shukla et al. 2010). In the promoter of the *Arabidopsis DNMT1*-homolog *MET1*, an AACGTAA sequence motif exists as well, although no report has yet proven that it could be targeted by *Arabidopsis* BRCA1 protein. There is no AACGTAA context in *DRM1* upstream sequence, but a truncated motif ACGTAA is observed. The interesting observation of X-ray induced upregulation was that transcripts of *BRCA1* and *DRM1* were both increased but not the transcript of MET1. It shows a possibility that *BRCA1* not only regulates *MET1* but also *DRM1*, when induced by DNA damage, through binding to AACGTAA/ACGTAA motif in *Arabidopsis*.

No previous report mentioned that *DRM1* and *DRM2* play roles in DDR. But the observed phenotype strongly suggests that *DRM1* is associated with X-ray induced DSBR. Although *de novo* methylation is directed by *DRM2* in *Arabidopsis*, *DRM1* possibly regulates the transcriptions of TEs and lncRNAs in response to stress, such as ionizing radiation, and this process is supposedly downstream of *BRCA1* (Figure 45). MET1 and HDA6 cooperate to inhibit TEs in *Arabidopsis* (Liu et al. 2012). They might negatively regulate lncRNAs generating from TEs/repeats loci as well. DRM1 probably has a similar function as MET1 to interact with histone modification proteins and regulates further processes of lncRNAs.

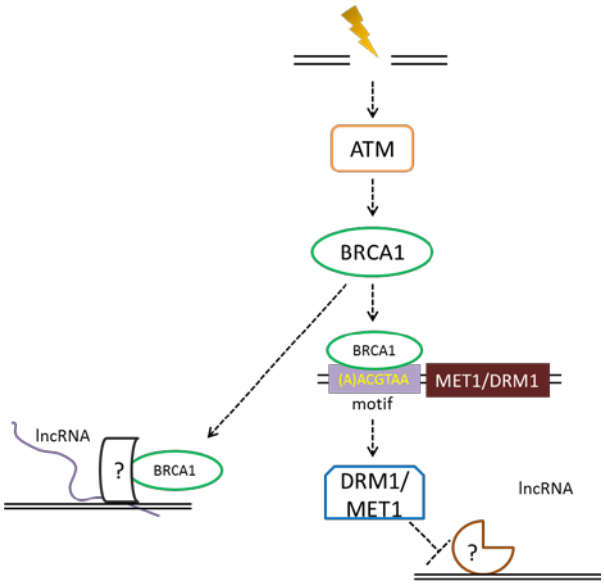


Figure 45: Hypothetical pathway of BRCA1 regulating MET1 and DRM1 in lncRNAs biosynthesis

ATM regulates BRCA1 in response to damage repair when DSBs initiated by X-ray. Transcription of *MET1* and *DRM1* is controlled by BRCA1, thus regulates future lncRNAs biosynthesis.

Though *drm1* and *brca1* mutants had similar X-ray induced phenotypes and regulation

patterns of partially detected transcripts in this work, it is still an open question how *DRM1* acts in that process, and whether it has more functions in response to DNA damage. For further analysis, an electrophoretic mobility shift assay (EMSA) can be used to identify the interaction of BRCA1 protein and promoters of *DRM1* and *MET1* in *Arabidopsis*. Double mutant of *brca1drm1* is also a good choice for following X-ray treatment, in order to investigate whether *BRCA1* and *DRM1* work in the same pathway.

4.2.6 Histone modifications were not generally altered 3 hours after X-ray irradiation

Histone modification is associated with altering gene expression in response to stresses (Luo et al. 2012). Methylation at histone 3 lysine 4 and lysine 36, and acetylation at specific loci on histone are associated with gene activation, whereas methylation of H3K9 and H3K27 plus H3 deacetylation are considered to be hallmarks of gene suppression (Liu et al. 2010a, Luo et al. 2012).

Histone methylation and demethylation are performed by corresponding methyltransferases (HKMTs) and demethylases (HDMs). Five subgroups of HKMTs catalyze mono-, di-, or trimethylation of amino acids in histone N-termini at different positions. For example, HKMT I members MEA, CLF, and SWN generate trimethylation on H3K27, while HKMT IV members ATXR5 and ATXR6 catalyze monomethylation of H3K27 in *Arabidopsis* (Thorstensen et al. 2011). Lysine Specific Demethylase 1 (*LSD1*) and JmjC domain-containing proteins are two families being responsible for histone demethylation. There are four LSD1-like coding genes in *Arabidopsis*, which remove methyl groups from H3K4 (Chen et al. 2011). A much bigger family is the JmjC domain-containing protein family with 21 JmjC containing genes in *Arabidopsis* (Lu et al. 2008). Their histone demethylation activities are only partially proven experimentally. For instance, JMJ11 (also named ELF6, early flowering 6) and JMJ25 (also named IBM1, increase in bonsai methylation 1) target H3K9 methylation (Saze et al. 2008, Yu et al. 2008). JMJ11 is also a H3K4 demethylase (Jeong et al. 2009), and JMJ12 (also named REF6, RELATIVE OF EARLY FLOWERING 6) is a H3K27 demethylase (Lu et al.

2011a).

Though there are a couple of genes involved in histone methylation and demethylation, surprisingly, only three of them were transcriptionally regulated in the early response to DNA damage. *ATXR6* and *ASHR3* were both downregulated. ATXR6 affects H3K27me1 and plays a role in silencing TEs and regulating DNA replication (Stroud et al. 2012). ASHR3 methylates both H3K4 and H3K36, and is involved in the regulation of stamen development (Cartagena et al. 2008, Thorstensen et al. 2008). JMJ30 is a circadian regulator in plants and humans. However, its enzymatic function of dimethylating H3K36 has not yet been proven in *Arabidopsis*. It was the only upregulated histone methylation-related gene 3 hours post irradiation in the WT plants.

There are four classes of histone acetyltransferases (HATs) in *Arabidopsis* with 12 members in total. 18 genes which can be categorized into three groups are predicted to encode histone deacetylases (HDACs) in *Arabidopsis* (Pandey et al. 2002). However, none of them were regulated after radiation. Only one Tudor domain and Acyl-CoA N-acyltransferase domain-containing gene AT5G58610 was upregulated 24 fold. Tudor domain recognizes methylated histones, Acyl-CoA N-acyltransferase acetylates histone. AT5G58610 is a putative defense gene (Libault et al. 2007, Ascencio-Ibanez et al. 2008), but its function in histone modification is still unclear. It is worth applying further experiments to unravel its potential role in histone acetylation, or even affecting histone methylation by induction of DSBs.

The anti-sera that were available in this work, anti-H3K4me2, anti-H3K4me3, anti-H3K36me2, and anti-H3K9ac, did not detect any obvious quantitative or qualitative alteration in modified histones after induction of DNA damage. This finding is consistent with the unchanged transcription levels of most histone modification genes. Taken together, it is concluded that 3 hours after induction of DSBs by X-ray treatment, histone modifications are either not y*et al*tered or alterations are below the detection limit. Either the DNA damage response in *Arabidopsis* lacks effects at the epigenetic chromatin status level (which is very unlikely), histone modifications are established only much later than expression responses of DDR proteins, or histone modifications occur which cannot be detected by antisera used (alternative amino acid positions, acetylation, etc.).

4.2.7 Does *JMJ30* participate in DSB-induced RdDM?

The Jumonji C (JmjC) domain is a signature domain of active histone demethylases which occurs widespread in yeast, humans and plants (Tsukada et al. 2006, Lu et al. 2008). In *Arabidopsis*, 21 JmjC domain-containing proteins are predicted (Lu et al. 2008). Identifying functions of JmjC proteins is still underway. So far, it is known about the *Arabidopsis* proteins that JMJ14 and JMJ15 are H3K4 demethylases, JMJ25 (IBM1) demethylates H3K9 (Chen et al. 2011), and JMJ12 (REF6) is a H3K27 demethylase (Lu et al. 2011a).

JMJ30 is the *Arabidopsis* ortholog of human JMJD5, which regulates the circadian system in both plants and humans (Jones et al. 2010). JMJD5 removes the two methyl groups from H3K36me2 and is associated with cell cycle progression in humans (Hsia et al. 2010). However, the enzymatic function of JMJ30 has not yet been confirmed in *Arabidopsis*.

In this study, irradiation retarded the growth of WT plants, whereas, in contrast, JMJ30 defective plants were more tolerant to X-ray. In human cells, dimethylation of H3K36 enhances DSB repair, mainly NHEJ, by promoting the association of early DNA repair components (Fnu et al. 2011). The upregulation of JMJ30 indicated a decreased H3K36me2 level in WT plants after damage. Consequently, in *jmj30* mutants the H3K36me2 level, and thus the NHEJ repair activity supposedly remain high after X-ray treatment, which may explain the radiation-resistant phenotype of the JMJ30-deficient plants (Figure 38). In future work, it will be interesting to study the progeny of the irradiated *jmj30* mutants to investigate whether more mutations were accumulated as a consequence of more NHEJ-repaired DSB sites.

JMJ14 is an active demethylase for H3K4, and it regulates flowering time of *Arabidopsis* (Lu et al. 2010). It is involved in DRM2-mediated RdDM to maintain non-CG methylation but is not required to establish *de novo* methylation (Deleris et al. 2010). The observation of JMJ14 regulated non-CG methylation via RdDM suggested another level of regulating DRM2 by histone (Deleris et al. 2010) rather than by AGO4-siRNA complex. H3K4 and H3K36 methylation are both active markers for transcription.

Therefore, it seems possible that the H3K36me2 demethylase JMJ30 might play a comparable role and participate in RdDM similarly as JMJ14.

4.2.8 Multiple functions of AGOs in response to DSBs?

In *Arabidopsis*, 10 Argonaute proteins are categorized in three subgroups based on their sequence similarity: AGO1/AGO5/AGO10 subgroup, AGO2/AGO3/AGO7 subgroup, and AGO4/AGO6/AGO8/AGO9 subgroup (Vaucheret 2008). AGOs are believed to act in plant defence against virus infection by RNA-directed gene silencing (Seo et al. 2013). They interact with small RNAs (sRNAs), but with length bias. AGO1, AGO2, and AGO7 prefer to bind 21 to 22 nt sRNAs, while AGO4, AGO6, and AGO9 prefer to load 24 to 26 nt sRNAs (Seo et al. 2013). AGO2 interacting sRNAs are mainly trans-acting siRNAs (ta-siRNAs) and repeat-associated siRNAs, whereas AGO4 and AGO6 load heterochromatic siRNAs (Meister 2013). In addition to RNA-directed gene silencing, AGO2 is the only known Argonaute protein that facilitates DSBR by loading sRNAs originating from proximal region of breaks (Wei et al. 2012). It also regulates Pol II activity (Cernilogar et al. 2011). In this research, AGO2 was upregulated three hours after irradiation in an ATM-dependent manner. Interestingly, *ago2* mutant plants were more tolerant than WT plants against X-ray irradiation, which is unexpected. The phenotype seems to conflict with the DSBR function of AGO2, which suggests other AGO proteins with redundant function in repair, and AGO2 rather regulates DSBs responsive genes negatively, if at all.

Another AGO protein regulated by X-ray in WT plants is AGO7, but it was not upregulated as strongly as AGO2. It belongs to the subgroup of AGO2, which means it might promote DSBR as well as AGO2. However, the qRT-PCR in *ago2* mutant did not show upregulation of *AGO7* three hours post-radiation. The *ago7* mutant acted similarly to WT after irradiation. So *AGO7* is not likey to play the same role in response to DSBs as *AGO2*.

The *ago4* mutant was tolerant to X-ray like the *ago2* mutant. But *AGO4* was transcripitionally unchanged in response to irradiation. *AGO4* is the key player of plant

specific small RdDM, and suppresses transposons in plants (Zilberman et al. 2003). In *AGO2* and *AGO4* defective plants, transcriptionally or post-transcriptionally silenced genes will have a higher transcriptional level than in WT. DDR related transcripts might also be among these upregulated ones, thus resulting in a higher tolerance to irradiation. AT1TE69975 and AT1TE04720 are two examples of AGO2 suppressed transcripts. They showed a higher transcriptional level in *ago2* mutant plants before X-ray treatment (see 3.3.4).

In addition to *AGO2*, *AGO4* and *AGO7*, other *AGOs* were not investigated in this work. Therefore, if other *AGOs* with redundant DNA damage reparing function as *AGO2* exist and are involved in the X-ray response is still elusive. QRT-PCR detected diversed transcriptional alterations of TEs and lncRNAs in *ago2*, *ago4* and *ago7* mutant plants, which suggests various regulations of *AGOs* after irradiation.

4.2.9 MicroRNA genes

In addition to protein coding genes and lncRNAs, two interesting microRNA genes were found to be regulated after X-ray irradiation. MiRNA genes MIR172 (AT2G28056) were reduced approximately 3-fold, and the other, MIR171C (AT1G62035) decreased more than 8-fold in irradiated WT plants.

MIR172 is a well establised microRNA, which regulates *APETALA2 (AP2)* gene and *TARGET OF EAT 3 (TOE3)*, restricts AGAMOUS (AG), and controls floral patterning in *Arabidopsis* (Jung et al. 2014, Spanudakis and Jackson 2014). By contrast to MIR172, there are only two articles available for MIR171C. It was reported that shoot branching can be regulated by targeting phytohormone GRAS family genes, *SCL6-II*, *SCL6-III*, and *SCL6-IV* in *Arabidopsis* (Wang et al. 2010). Its homolog plays a role in the induction process of larch Somatic embryogenesis (Zhang et al. 2012).

In this study, upregulation of *AP2* genes probably resulted from the decreased expression of MIR172, in order to respond to the DNA damage induced hormone alteration. In WT plants, repressed MIR171C caused a slight upregulation of *SCL6-IV*. However, MIR171C was downregulated in the *atm* mutant.

The MIRNA genes detected by RNA-seq expand the range of transcriptomic response to X-ray induced DSBs in *Arabidopsis*.

4.2.10 Model of TEs/repeats in response to DSBs

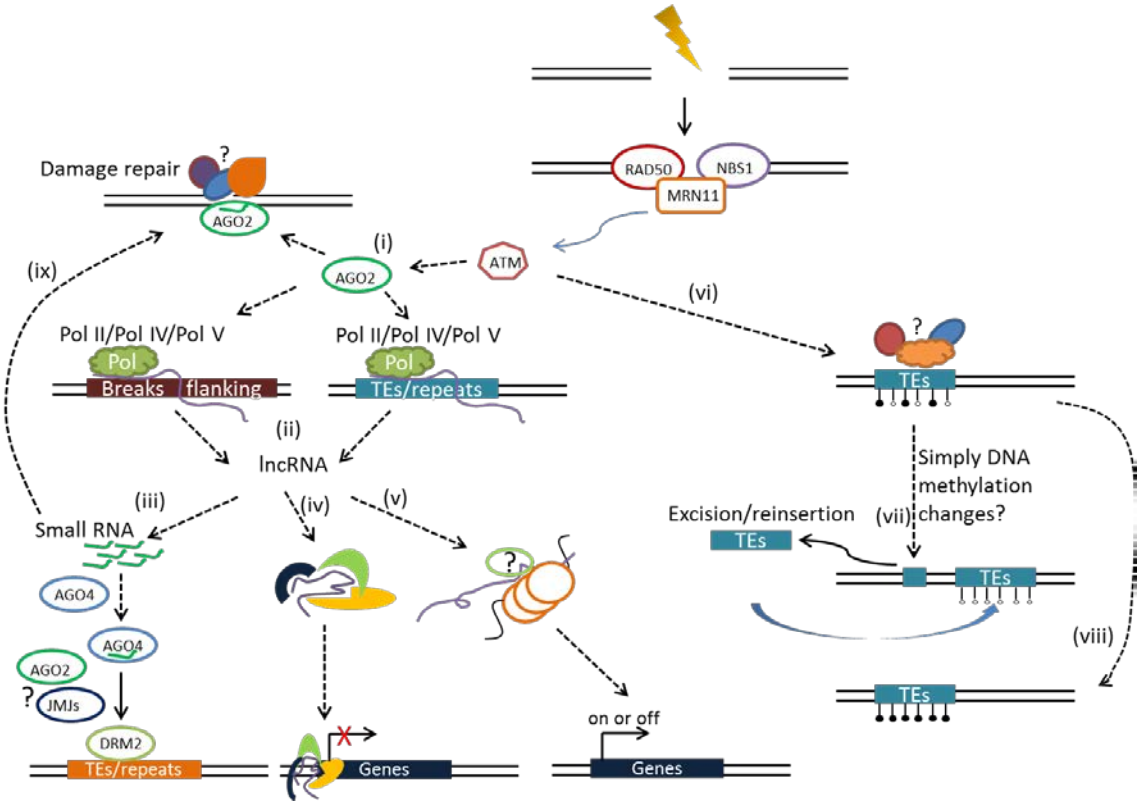


Figure 46: Schematic illustration of potential responses of TEs/repeats to DSBs

Combining the data generated in this study with previous reports, a model is suggested for the differential responses of TEs after DSB induction in *Arabidopsis* (Figure 46). When DNA damage occurs, ATM initiates downstream responses (Garcia et al. 2003, Uziel et al. 2003). It may control the biosynthesis of lncRNAs from breaks in flanking regions or TEs/repeats regions by Pol II, Pol IV or Pol V via AGO2 (i and ii) (Cernilogar et al. 2011, Wei et al. 2012). LncRNAs can function in three different ways. The first is to generate small RNAs and participate in RdDM associated with AGO4, DRM2, probably AGO2 (Wierzbicki et al. 2009, Wei et al. 2012) and also JMJ proteins (iii). The second is to serve as a scaffold to recruit protein complexes and target them to specific loci (Wierzbicki 2012), thereby further suppressing the target genes (iv). The third could be

to bind to chromatin modifiers and switch on or off the gene (v) (De Lucia and Dean 2011). In addition to regulating RNA polymerases and recruiting small RNAs in order to target specific genes or loci, AGO2 can load small RNAs derived from flanking DNA breaks, and then facilitate the process of DSBR (ix) (Wei et al. 2012). Of course this model does not exclude the possibility that certain TEs are regulated by other pathways (vi), such as simply by DNA methylation changes inducing activation or suppression (vii and viii).

4.3 *Ac*, an example of defending heterologous TEs in *Arabidopsis*?

In one of the earliest attempts to introduce the maize transposable element *Ac* into *Arabidopsis*, Schmidt and Willmitzer observed that germinal excision frequency was as low as 0.2 - 0.5 % in transgenic *Arabidopsis* (Schmidt and Willmitzer 1989). In another study, germinal excision frequencies of *Ac* ranging from 0.07 % to 5.7 % in *Arabidopsis* were observed (Dean and Lister 1992). However, the germinal excision frequency of *Ac* was up to 50 % in tobacco (Kunze et al. 1995).

The reason of low germinal transposition frequency in *Arabidopsis* is still unclear. But the observed mis-processed transcripts of *AcTPase* in *Arabidopsis* (Jarvis et al. 1997) and sugar beet (Lisson et al. 2010) could be an explanation. The mis-processed *Ac* transcripts might diminish transposition efficiency via post-transcriptional, translational and post-translational regulation. Truncated *Ac* transcripts could be a source for generating small RNAs leading to degradation of correct *Ac* transcript, therefore, results in decreased AcTPase amount. If truncated *Ac* transcripts are still translated, but in unfunctional formation, the function-lost AcTPases might block interactive location of *Ac/Ds* in competition with full-functional AcTPase. These possibilities all lead to the decreased transposition efficiency in *Arabidopsis*.

It is possible to increase transposition frequency by preventing incorrect *AcTPase* transcripts. This can be achieved by mutating the reported alternative splice sites or early termination positions without changing the amino acid sequence of the *Ac*

transposase protein. The results of semi-quantitative RT-PCR in this study showed that mis-processed transcripts in some degree disappeared in plants containing modified *AcTPase* CDSs, M3 and M7 having 3 and 7 mutations, respectively. But the excision frequencies were decreased in *AcTPase* mutant and *Ds* crossed plants. It seems that the point mutations could block those cryptic splicing sites, therefore no multiple bands were observed in the PCR products of mutated plants. The amount of *AcTPase* transcripts in some of M3_14 and M7_1 plants were less than in WT plants. In maize and tobacco, a negative dosage effect of *Ac* transposition was observed (McClintock 1949, McClintock 1951, Scofield et al. 1993). In order to detect if the amount of transcripts in M3 and M7 plants generated a higher amount of TPase, Western blot was carried out. However, no TPase protein was detectable, neither in WT or mutant *AcTPase* plants. This suggests that the decrease of transposition was not due to the negative dosage effect of *AcTPase* in mutant plants. It is more likely that just the increased transcripts caused the higher transposition efficiency in WT *Ac* plants.

Northern blot analysis did not detect mis-processed *Ac* transcripts in maize and transgenic tobacco (Kunze and Stochaj U 1987, Hehl and Baker 1990). However, in a later work, incorrect splicing of *Ac* transcripts was observed in tobacco, sugar beet and *Arabidopsis (Jarvis et al. 1997, Lisson et al. 2010)*. But the alternative splicing was much less frequent in tobacco than in sugar beet (Lisson et al. 2010). Since the incorrect transcripts of *Ac* occurred not only in *Arabidopsis,* but also in sugar beet and tobacco, this could indicate a possible common defense mechanism of plants against heterologous active transposon. As a transposase CDS, heterologous plants probably recognize deviating codon usage or GC-content and in turn generate incorrect transcripts, and further those mis-processed transcripts could lead to RNA degradation via generating small RNA. Finally, the transposition of the transposable element will be blocked or decreased. Although it seems the excision efficiencies in tobacco and sugar beet were not really affected by this machinery (Baker et al. 1986, Finnegan et al. 1988, Lisson et al. 2010), this might be an explanation for the low germinal excision in *Arabidopsis*.

Comparing with Maize, *Arabidopsis* owns a smaller genome size and only 17 % TEs in

genomic DNA. It probably indicates less tolerance of TEs in *Arabidopsis* than in Maize. When heterologous TEs are transformed into *Arabidopsis*, there might be an anti-machinery to suppress them. If there is a defense process to deal with specific heterologous DNA transformed in *Arabidopsis*, then it is not promising to prevent the mis-processed *Ac* transcripts by mutating only some of those cryptic termination sites or alternative splice sites. In transgenic *Arabidopsis* containing WT *Ac* coding sequence controlled by 35S promoter, there were more early terminated transcripts observed. Those early terminated positions all include AATAAA or a similar signal region of termination in the Ac sequence. It is interesting to note that in maize they are not recognized as polyadenylation sites but in *Arabidopsis*, perhaps in other heterologous plants transformed with *Ac*, those sites were terminated earlier as well. In further work, *Ac* transcripts could be investigated in *Arabidopsis* and maize mutant plants lacking transcription-controlling genes, such as RNA polymerase genes. This might be helpful to know whether the differing transcriptional machinery causes abnormal *Ac* transcripts in *Arabidopsis*.

5 Summary

In this study two major topics were investigated. The first part deals with the early genomic response to DNA damage with particular emphasis on transposable elements (TEs). The second part aims to improve *Ac* transposition by suppressing the formation of mis-processed *Ac* transcripts by introducing point mutations in the *Ac* transposase coding sequence.

1) DNA damage and TEs

Double-strand breaks (DSBs) can lead to genome instability and transcriptional and transpositional reactivation of TEs. However the mechanism underlying this process is not well understood.

Two different strategies were applied to generate DSBs and to follow transposon reactivation activity. Initially four coding sequences of meganuclease (*ISI-opA*, *I-Sce I*, *ZFN3* and *QQR*), under the control of an estradiol-inducible promoter, and their respective target sites were introduced into the genome of *Arabidopsis*. Only the I-Sce I endonuclease, encoded by the *Arabidopsis* codon-optimized sequence *ISI-opA*, successfully generated DSBs after estradiol induction. However, the expected downstream response to DSBs was not observed probably due to a low frequency of introduced breaks into the genome.

As an alternative strategy, X-rays were used to generate DSBs in the *Arabidopsis* genome. To compare the transcriptome-wide early response to DSBs in WT and the DNA-repair-defective mutant *atm*, the RNA-seq technology was used as a tool. 1315 and 644 genes were found to be e2-fold regulated in WT and *atm* respectively, largely consistent with previous microarray-based transcriptome studies. However, the RNA-seq technology allowed obtaining additional information concerning previously undetected events in the genome response.

RNA-Seq data revealed that, in contrast to the large number of regulated genes, TEs/TE-related elements were less responsive on transcriptional level to DSBs, at least in the early response (3h after the treatment). The DSBs induced transcriptional activation of TEs/TE-related elements observed in this study is consistent with previous published data. However, in this work DNA damage induced transcriptional down-regulation of TEs/TE-related elements was detected at transcriptome level for the

first time. Among the regulated TEs, retrotransposons were more responsive to X-ray than DNA transposons. In WT plants, 68 % of the regulated TEs/TE-related elements were associated with long non-coding RNAs (lncRNAs). In addition, several differentially expressed novel transcripts were detected and were also identified as lncRNAs. In total, 91 regulated lncRNAs were identified in WT, but only 18 in the *atm* mutant. The observation indicated that lncRNAs were regulated by ATM in response to X-ray induced DSBs in *Arabidopsis*.

The epigenetic machinery is associated with DNA damage repair and the regulation of TEs. In this study, the *de novo* DNA methylation gene *DRM1* was the only up-regulated DNA methyltransferase gene induced by X-ray. The demethylase gene *JMJ30* and two RNA-directed DNA methylation (RdDM) genes *AGO2* and *AGO7* were also upregulated. In order to further investigate the roles of these epigenetic-related genes in regulating TEs/TE-related elements and lncRNAs in response to DSBs, the *Arabidopsis* T-DNA insertion knock-out lines *drm1*, *jmj30*, *ago2* and *ago7* mutant were X-ray treated. BRCA1 repairs DSBs via homologous recombination. DRM2, the homolog of DRM1, is the mainly active methyltransferase for *de novo* DNA methylation and AGO4 is the key player in the RdDM pathway. The corresponding T-DNA insertion lines were also irradiated. The phenotypical analysis of mutants and the quantitative expression analysis of a subset of TEs/TE-related elements and lncRNAs via qRT-PCR indicated that DNA damage repair genes and genes involved in epigenetic control might selectively regulate TEs/TE-related elements and lncRNAs in response to DSBs.

2) Improvement of *Ac* transposition in *Arabidopsis*

Previous studies have reported that the expression of *Ac* in *Arabidopsis* results in a low germinal excision rate which might be due to the occurrence of cryptic introns and early terminated *Ac* transcripts. Site-directed mutagenesis of *Ac* was used in the present study in order to prevent mis-processing and incorrect splicing of *Ac* transcripts and to improve *Ac* transposition frequency. However, the modified *AcTPase* coding sequence did not increase transposition frequency. Additional studies are needed to determine if the different transcription-controlling mechanisms in *Arabidopsis* and maize might be responsible for the occurrence of mis-processed *Ac* transcripts in *Arabidopsis.*

6 Zusammenfassung

In dieser Arbeit wurden zwei Hauptfragestellungen bearbeitet. Im ersten Teil wurde die schnelle genomische Antwort auf DNA Schäden untersucht, wobei der Schwerpunkt auf die transponierende Elemente (TEs) gelegt wurde. Ziel des zweiten Teils der Arbeit war es, die Transposition des *Ac* Transposons zu verbessern. Dazu wurden Punktmutationen in die kodierende Sequenz des *Ac* Transposons eingefügt, um die Anzahl an mis-prozessierten *Ac* Transkripten zu verringern.

1) DNA Schädigung und TEs

DNA-Doppelstrangbrüche (DSBs) können die Integrität des Genoms negativ beeinflussen. Eine der Folgen kann die transkriptionelle und transpositionelle Reaktivierung von Transposons sein. Das Verständnis der zugrunde liegenden Mechanismen hinter dieser Beobachtung ist jedoch limitiert.

Es wurden zwei Strategien angewandt, um DSBs zu erzeugen und die Reaktivierung der Transposonaktivität zu untersuchen. Zunächst wurden die kodierenden Sequenzen der vier Meganukleasen *ISI-opA*, *I-Sce I*, *ZFN3* und *QQR* und ihre Erkennungssequenzen in *Arabidopsis* transformiert. Für die Kontrolle der Expression der Meganukleasen wurde ein Östradiol-induzierbares System verwendet. Lediglich nach der Expression der *I-Sce I*-Endonuklease, die durch die *Arabidopsis*-optimierte kodierende Sequenz *ISI-opA* kodiert wurde, konnten DSBs erfolgreich generiert werden. Allerdings konnten die erwarteten Antworten auf die DSBs im Genom nicht beobachtet werden, was vermutlich auf eine zu geringe Anzahl an DBSs zurückzuführen war.

Alternativ wurden *Arabidopsis*-Pflanzen Röntgenstrahlen ausgesetzt, um DSBs zu generieren. Zur umfassenden Analyse von frühzeitigen, transkriptionellen Reaktionen auf solche Ereignisse im Wildtyp und in der im DNA Reparaturmechanismus defizienten Mutante *atm* wurde die Technologie der RNA-Sequenzierung (RNA-seq) verwendet. Dabei stellte sich heraus, dass 1315 Gene in Wildtyp-Pflanzen und 644 Gene in der *atm*-Mutante mehr als zweifach reguliert waren. Die erhaltenen Daten entsprachen den Ergebnissen vorheriger Microarray-Untersuchungen zu DNA-Schäden in *Arabidopsis*, wobei die Verwendung der RNA-Seq Technologie es ermöglichte, neue Informationen über bislang unbekannte Vorgänge bzw. Mechanismen der schnellen Genomantwort zu erhalten.

So konnte gezeigt werden, dass TEs und TE-ähnliche Elemente im Gegensatz zu den meisten regulierten Genen weniger auf DSBs reagieren (bezogen auf die Analyse 3h nach der Röntgenbehandlung). Die durch DSBs induzierte Aktivierung von TEs und TE-ähnlichen Elementen ist konsistent mit bereits publizierten Artikeln. Allerdings konnte nun erstmalig

auch eine Herunterregulierung von TEs und TE-ähnliche Elementen durch Schädigungen der DNA auf Transkriptionsebene beobachtet werden. Unter allen regulierten TEs, reagierten Retrotransposons stärker auf Röntgenbestrahlung als DNA Transposons. In Wildtyp-Pflanzen waren 68 % der regulierten TEs und TE-ähnliche Elemente mit sogenannten langen-nicht-kodierenden RNAs (*long non-coding* RNAs, lncRNAs) assoziiert. Weitere, bisher unbekannte und differentiell exprimierte Transkripte konnten ebenfalls als lncRNAs identifiziert werden. Insgesamt wurden 91 lncRNAs in Wildtyp-Pflanzen reguliert, aber lediglich 18 in *atm*-Mutanten. Diese Beobachtung wies darauf hin, dass lncRNAs als Antwort auf Röntgenstrahlen-induzierte DSBs in Arabidopsis durch ATM reguliert werden.
Die epigenetische Maschinerie ist mit der Reparatur von DNA Schäden und der Regulierung von TEs assoziiert. In dieser Arbeit konnte gezeigt werden, dass *DRM1* als einzige DNA Methyltransferase durch Röntgenstrahlen in seiner Genexpression hoch-reguliert wurde. Das Gen der Demethylase *JMJ30* als auch die Gene *AGO2* und *AGO7*, die in der RNA-vermittelten DNA-Methylierung (RdDM) involviert sind, wurden ebenfalls in ihrer Expression hoch-reguliert. Um den regulatorischen Einfluss dieser Gene auf TEs/TE-ähnlichen Elementen und lncRNAs zu analysieren, wurden die *Arabidopsis* T-DNA Insertionslinien *drm1*, *jmj30*, *ago2* und *ago7* mit Röntgenstrahlen behandelt. BRCA1 ist an der Reparatur von Doppelstrangbrüchen der DNA mittels homologer Rekombination beteiligt. DRM2, ein Homolog von DRM1 ist die wichtigste Methyltransferase für die *de novo* DNA Methylierung und AGO4 ist eine Schlüsselkomponente im RdDM Weg. T-DNA Insertionslinien dieser Mutanten wurden ebenfalls mit Röntgenstrahlen behandelt. Die phänotypische Analyse der Mutanten, sowie die Expressionsanalyse ausgewählter TEs/TE-ähnlicher Elemente und von lncRNAs mittels qRT-PCR deuten darauf hin, dass Gene mit Funktion in der DNA-Reparatur und der epigenetischer Kontrolle von TEs/TE-ähnlichen Elementen und lncRNAs gezielt durch DSBs reguliert wurden.

2) Verbesserung der *Ac* Transposition in *Arabidopsis*

In früheren Studien wurde gezeigt, dass die Expression von *Ac* in *Arabidopsis* zu einer verringerten germinalen Exzisionsrate führt. Dies könnte eine Folge von kryptischen Introns und vorzeitig abgebrochenen Transkripten sein. Um eine Misprozessierung und ein inkorrektes Spleißen von Ac Transkripten zu reduzieren und somit die Transpositionsfrequenz von Ac zu erhöhen, wurde eine zielgerichtete Mutagenese der kodierenden Sequenz von *Ac* durchgeführt. Jedoch führte dieser Versuchsansatz zu keiner Erhöhung der Transpositionsfrequenz in *Arabidopsis*.

7 References

ABRAHAM RT. 2003. Checkpoint signaling: epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase. Bioessays 25: 627-630.

ADAMS MD ET AL. 2000. The genome sequence of Drosophila melanogaster. Science 287: 2185-2195.

AFA S, HUBLEY R AND GREEN P. 1996-2014. Repeatmasker open-4.0.5.

AGIUS F, KAPOOR A AND ZHU JK. 2006. Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. Proceedings of the National Academy of Sciences of the United States of America 103: 11796-11801.

ALTMANN T, SCHMIDT R AND WILLMITZER L. 1992. Establishment of a Gene Tagging System in Arabidopsis-Thaliana Based on the Maize Transposable Element Ac. Theoretical and Applied Genetics 84: 371-383.

ARIEL F, JEGU T, LATRASSE D, ROMERO-BARRIOS N, CHRIST A, BENHAMED M AND CRESPI M. 2014. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. Molecular cell 55: 383-396.

ARNOULD S, DELENDA C, GRIZOT S, DESSEAUX C, PAQUES F, SILVA GH AND SMITH J. 2011. The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel 24: 27-31.

ARVIDSSON S, KWASNIEWSKI M, RIANO-PACHON DM AND MUELLER-ROEBER B. 2008. QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC bioinformatics 9: 465.

ASCENCIO-IBANEZ JT, SOZZANI R, LEE TJ, CHU TM, WOLFINGER RD, CELLA R AND HANLEY-BOWDOIN L. 2008. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant physiology 148: 436-454.

ATHMA P, GROTEWOLD E AND PETERSON T. 1992. Insertional mutagenesis of the maize P gene by intragenic transposition of Ac. Genetics 131: 199-209.

AYMARD F ET AL. 2014. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nature structural & molecular biology 21: 366-374.

BAKER B, COUPLAND G, FEDOROFF N, STARLINGER P AND SCHELL J. 1987. Phenotypic assay for excision of the maize controlling element Ac in tobacco. The EMBO journal 6: 1547-1554.

BAKER B, SCHELL J, LORZ H AND FEDOROFF N. 1986. Transposition of the maize controlling element "Activator" in tobacco. Proceedings of the National Academy of Sciences of the United States of America 83: 4844-4848.

BANKS JA, MASSON P AND FEDOROFF N. 1988. Molecular mechanisms in the developmental regulation of the maize Suppressor-mutator transposable element. Genes & development 2: 1364-1380.

BAUCOM RS, ESTILL JC, CHAPARRO C, UPSHAW N, JOGI A, DERAGON JM, WESTERMAN RP, SANMIGUEL PJ AND BENNETZEN JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS genetics 5: e1000732.

BELFORT M AND ROBERTS RJ. 1997. Homing endonucleases: keeping the house in order. Nucleic acids research 25: 3379-3388.

BENNETZEN JL AND WANG H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annual review of plant biology 65: 505-530.

BERR A, SHAFIQ S AND SHEN WH. 2011. Histone modifications in transcriptional activation during plant development. Biochimica et biophysica acta 1809: 567-576.

BERR A, XU L, GAO J, COGNAT V, STEINMETZ A, DONG A AND SHEN WH. 2009. SET DOMAIN GROUP25 encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering. Plant physiology 151: 1476-1485.

BIBIKOVA M, CARROLL D, SEGAL DJ, TRAUTMAN JK, SMITH J, KIM YG AND CHANDRASEGARAN S. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Molecular and cellular biology 21: 289-297.

BIRD A, TATE P, NAN X, CAMPOY J, MEEHAN R, CROSS S, TWEEDIE S, CHARLTON J AND MACLEOD D. 1995. Studies of DNA methylation in animals. Journal of cell science Supplement 19: 37-39.

BLEIN JP, COUTOS-THEVENOT P, MARION D AND PONCHET M. 2002. From elicitins to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. Trends in plant science 7: 293-296.

BLEWITT ME, VICKARYOUS NK, PALDI A, KOSEKI H AND WHITELAW E. 2006. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS genetics 2: e49.

BOEHM U, HEINLEIN M, BEHRENS U AND KUNZE R. 1995. One of three nuclear localization signals of maize Activator (Ac) transposase overlaps the DNA-binding domain. Plant Journal 7: 441-451.

BOSTELMAN LJ, KELLER AM, ALBRECHT AM, ARAT A AND THOMPSON JS. 2007. Methylation of histone H3 lysine-79 by Dot1p plays multiple roles in the response to UV damage in Saccharomyces cerevisiae. DNA repair 6: 383-395.

BOYKO A, ZEMP F, FILKOWSKI J AND KOVALCHUK I. 2006. Double-strand break repair in plants is developmentally regulated. Plant physiology 141: 488-497.

BUCHER E, REINDERS J AND MIROUZE M. 2012. Epigenetic control of transposon transcription and mobility in Arabidopsis. Current opinion in plant biology 15: 503-510.

BUISINE N, QUESNEVILLE H AND COLOT V. 2008. Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets. Genomics 91: 467-475.

BUNTING SF ET AL. 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell 141: 243-254.

CAMPI M, D'ANDREA L, EMILIANI J AND CASATI P. 2012. Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. Plant physiology 158: 981-995.

CAO X, AUFSATZ W, ZILBERMAN D, METTE MF, HUANG MS, MATZKE M AND JACOBSEN SE. 2003. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Current biology : CB 13: 2212-2217.

CAO X AND JACOBSEN SE. 2002. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current biology : CB 12: 1138-1144.

CAO X, SPRINGER NM, MUSZYNSKI MG, PHILLIPS RL, KAEPPLER S AND JACOBSEN SE. 2000. Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proceedings of the National Academy of Sciences of the United States of America 97: 4979-4984.

CARR M, BENSASSON D AND BERGMAN CM. 2012. Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. PloS one 7: e50978.

CARTAGENA JA, MATSUNAGA S, SEKI M, KURIHARA D, YOKOYAMA M, SHINOZAKI K, FUJIMOTO S, AZUMI Y, UCHIYAMA S AND FUKUI K. 2008. The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Developmental biology 315: 355-368.

CASATI P, CAMPI M, CHU F, SUZUKI N, MALTBY D, GUAN S, BURLINGAME AL AND WALBOT V. 2008. Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. The Plant cell 20: 827-842.

CERNILOGAR FM ET AL. 2011. Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature 480: 391-395.

CHARBONNEL C, ALLAIN E, GALLEGO ME AND WHITE CI. 2011. Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. DNA repair 10: 611-619.

CHAUDHURI S, WYRICK JJ AND SMERDON MJ. 2009. Histone H3 Lys79 methylation is required for efficient nucleotide excision repair in a silenced locus of Saccharomyces cerevisiae. Nucleic acids research 37: 1690-1700.

CHEN IP, MANNUSS A, OREL N, HEITZEBERG F AND PUCHTA H. 2008. A homolog of ScRAD5 is involved in DNA repair and homologous recombination in Arabidopsis. Plant physiology 146: 1786-1796.

CHEN L, SONG Y, LI S, ZHANG L, ZOU C AND YU D. 2012. The role of WRKY transcription factors in plant abiotic stresses. Biochimica et biophysica acta 1819: 120-128.

CHEN X, HU Y AND ZHOU DX. 2011. Epigenetic gene regulation by plant Jumonji group of histone demethylase. Biochimica et biophysica acta 1809: 421-426.

CHENG C, DAIGEN M AND HIROCHIKA H. 2006. Epigenetic regulation of the rice retrotransposon Tos17. Molecular genetics and genomics : MGG 276: 378-390.

CHEVALIER BS AND STODDARD BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic acids research 29: 3757-3774.

CLAY NK, ADIO AM, DENOUX C, JANDER G AND AUSUBEL FM. 2009. Glucosinolate metabolites required for an Arabidopsis innate immune response. Science 323: 95-101.

CLOUGH SJ AND BENT AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant Journal 16: 735-743.

CREASEY KM, ZHAI J, BORGES F, VAN EX F, REGULSKI M, MEYERS BC AND MARTIENSSEN RA. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature 508: 411-415.

CUI H AND FEDOROFF NV. 2002. Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. The Plant cell 14: 2883-2899.

CUI X ET AL. 2013. Control of transposon activity by a histone H3K4 demethylase in rice.

Proceedings of the National Academy of Sciences of the United States of America 110: 1953-1958.

CULLIGAN KM, ROBERTSON CE, FOREMAN J, DOERNER P AND BRITT AB. 2006. ATR and ATM play both distinct and additive roles in response to ionizing radiation. Plant Journal 48: 947-961.

CUOZZO C ET AL. 2007. DNA damage, homology-directed repair, and DNA methylation. PLoS genetics 3: e110.

CURTIS MD AND GROSSNIKLAUS U. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant physiology 133: 462-469.

DAUGAARD M ET AL. 2012. LEDGF (p75) promotes DNA-end resection and homologous recombination. Nature structural & molecular biology 19: 803-810.

DE LUCIA F AND DEAN C. 2011. Long non-coding RNAs and chromatin regulation. Current opinion in plant biology 14: 168-173.

DEAN C AND LISTER CSTPJJC. 1992. Behaviour of the maize transposable element Ac in Arabidopsis thaliana. The Plant J 2: 69-81.

DELERIS A, GREENBERG MV, AUSIN I, LAW RW, MOISSIARD G, SCHUBERT D AND JACOBSEN SE. 2010. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. EMBO reports 11: 950-955.

DESAI UJ AND PFAFFLE PK. 1995. Single-step purification of a thermostable DNA polymerase expressed in Escherichia coli. BioTechniques 19: 780-782, 784.

DING Y ET AL. 2007. SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. The Plant cell 19: 9-22.

DOLL M-L. 2008. Entwicklung eines Ds-Exzisionsassays mit einem bifunktionalen Reporter in stabil transformierten Arabidopsis thaliana Pflanzen. Diploma thesis.

DOUCET-CHABEAUD G, GODON C, BRUTESCO C, DE MURCIA G AND KAZMAIER M. 2001. Ionising radiation induces the expression of PARP-1 and PARP-2 genes in Arabidopsis. Molecular genetics and genomics : MGG 265: 954-963.

DRURY GE, DOWLE AA, ASHFORD DA, WATERWORTH WM, THOMAS J AND WEST CE. 2012. Dynamics of Plant Histone Modifications in Response to the DNA Damage. The Biochemical journal 445(3): 393-401 .

DU Z, ZHOU X, LING Y, ZHANG Z AND SU Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research 38: W64-70.

DUJON B. 1980. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. Cell 20: 185-197.

DUPRE A, BOYER-CHATENET L AND GAUTIER J. 2006. Two-step activation of ATM by DNA and the Mre11-Rad50-Nbs1 complex. Nature structural & molecular biology 13: 451-457.

EARLEY KW, SHOOK MS, BROWER-TOLAND B, HICKS L AND PIKAARD CS. 2007. In vitro specificities of Arabidopsis co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. The Plant journal : for cell and molecular biology 52: 615-626.

ELGAR G ET AL. 1999. Generation and analysis of 25 Mb of genomic DNA from the pufferfish Fugu rubripes by sequence scanning. Genome research 9: 960-971.

EMELYANOV A GY, NAQVI NI, PARINOV S. 2006. Trans-kingdom transposition of the maize

dissociation element. Genetics 174: 1095-1104.

ENGREITZ JM ET AL. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341: 1237973.

FARKASH EA, KAO GD, HORMAN SR AND PRAK ET. 2006. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. Nucleic acids research 34: 1196-1204.

FEDOROFF NV. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. Science 338: 758-767.

FENG S, JACOBSEN SE AND REIK W. 2010. Epigenetic reprogramming in plant and animal development. Science 330: 622-627.

FINNEGAN EJ, TAYLOR B, DENNIS E AND PEACOCK WJ. 1988. Transcription of the maize transposable element Ac in maize seedlings and in transgenic tobacco. Mol Gen Genet 212: 505-509.

FLADUNG M AND POLAK O. 2012. Ac/Ds-transposon activation tagging in poplar: a powerful tool for gene discovery. BMC genomics 13: 61.

FNU S, WILLIAMSON EA, DE HARO LP, BRENNEMAN M, WRAY J, SHAHEEN M, RADHAKRISHNAN K, LEE SH, NICKOLOFF JA AND HROMAS R. 2011. Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proceedings of the National Academy of Sciences of the United States of America 108: 540-545.

FROSCHAUER A, SPROTT D, GERWIEN F, HENKER Y, RUDOLPH F, PFENNIG F AND GUTZEIT HO. 2012. Effective generation of transgenic reporter and gene trap lines of the medaka (Oryzias latipes) using the Ac/Ds transposon system. Transgenic Res 21: 149-162.

GALLEGO ME, JEANNEAU M, GRANIER F, BOUCHEZ D, BECHTOLD N AND WHITE CI. 2001. Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity. The Plant journal : for cell and molecular biology 25: 31-41.

GARCIA V, BRUCHET H, CAMESCASSE D, GRANIER F, BOUCHEZ D AND TISSIER A. 2003. AtATM is essential for meiosis and the somatic response to DNA damage in plants. The Plant cell 15: 119-132.

GONG Z AND ZHU JK. 2011. Active DNA demethylation by oxidation and repair. Cell research 21: 1649-1651.

GOSPODINOV A AND HERCEG Z. 2013. Chromatin structure in double strand break repair. DNA repair 12: 800-810.

GOSPODINOV A, VAISSIERE T, KRASTEV DB, LEGUBE G, ANACHKOVA B AND HERCEG Z. 2011. Mammalian Ino80 mediates double-strand break repair through its role in DNA end strand resection. Molecular and cellular biology 31: 4735-4745.

GRANDBASTIEN MA. 2004. [Stress activation and genomic impact of plant retrotransposons]. J Soc Biol 198: 425-432.

GRANDBASTIEN MA, LUCAS H, MOREL JB, MHIRI C, VERNHETTES S AND CASACUBERTA JM. 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica 100: 241-252.

GRECO R, OUWERKERK PB, TAAL AJ, FAVALLI C, BEGUIRISTAIN T, PUIGDOMENECH P, COLOMBO L, HOGE JH AND PEREIRA A. 2001. Early and multiple Ac transpositions in rice suitable for efficient insertional mutagenesis. Plant molecular biology 46: 215-227.

GREVELDING C, BECKER D, KUNZE R, VON MENGES A, FANTES V, SCHELL J AND MASTERSON R. 1992. High rates of Ac/Ds germinal transposition in Arabidopsis suitable for gene isolation by insertional mutagenesis. Proceedings of the National Academy of Sciences of the United States of America 89: 6085-6089.

GRUNTMAN E, QI Y, SLOTKIN RK, ROEDER T, MARTIENSSEN RA AND SACHIDANANDAM R. 2008. Kismeth: analyzer of plant methylation states through bisulfite sequencing. BMC bioinformatics 9: 371.

HARVEY JJ, LEWSEY MG, PATEL K, WESTWOOD J, HEIMSTADT S, CARR JP AND BAULCOMBE DC. 2011. An antiviral defense role of AGO2 in plants. PloS one 6: e14639.

HASHIDA SN, KITAMURA K, MIKAMI T AND KISHIMA Y. 2003. Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in Antirrhinum majus. Plant physiology 132: 1207-1216.

HE XJ, CHEN TP AND ZHU JK. 2011. Regulation and function of DNA methylation in plants and animals. Cell research 21: 442-465.

HE XJ, HSU YF, ZHU S, LIU HL, PONTES O, ZHU J, CUI X, WANG CS AND ZHU JK. 2009. A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes & development 23: 2717-2722.

HEGDE ML, HAZRA TK AND MITRA S. 2008. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. Cell research 18: 27-47.

HEHL R AND BAKER B. 1990. Properties of the maize transposable element Activator in transgenic tobacco plants: a versatile inter-species genetic tool. The Plant cell 2: 709-721.

HETZL J, FOERSTER AM, RAIDL G AND MITTELSTEN SCHEID O. 2007. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. The Plant journal : for cell and molecular biology 51: 526-536.

HSIA DA, TEPPER CG, POCHAMPALLI MR, HSIA EY, IZUMIYA C, HUERTA SB, WRIGHT ME, CHEN HW, KUNG HJ AND IZUMIYA Y. 2010. KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. Proceedings of the National Academy of Sciences of the United States of America 107: 9671-9676.

HUDSON ME, LISCH DR AND QUAIL PH. 2003. The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. The Plant journal : for cell and molecular biology 34: 453-471.

HUEN MS, SY SM AND CHEN J. 2010. BRCA1 and its toolbox for the maintenance of genome integrity. Nature reviews Molecular cell biology 11: 138-148.

HUETTEL B, KANNO T, DAXINGER L, AUFSATZ W, MATZKE AJ AND MATZKE M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. The EMBO journal 25: 2828-2836.

HUYEN Y, ZGHEIB O, DITULLIO RA, JR., GORGOULIS VG, ZACHARATOS P, PETTY TJ, SHESTON EA, MELLERT HS, STAVRIDI ES AND HALAZONETIS TD. 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432: 406-411.

IIDA S, MORITA Y, CHOI JD, PARK KI AND HOSHINO A. 2004. Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. Adv Biophys 38: 141-159.

ISHIMURA A, MINEHATA K, TERASHIMA M, KONDOH G, HARA T AND SUZUKI T. 2012. Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. Development 139: 749-759.

IZAWA T, MIYAZAKI C, YAMAMOTO M, TERADA R, IIDA S AND SHIMAMOTO K. 1991. Introduction and transposition of the maize transposable element Ac in rice (Oryza sativa L.). Mol Gen Genet 227: 391-396.

JACKSON JP, JOHNSON L, JASENCAKOVA Z, ZHANG X, PEREZBURGOS L, SINGH PB, CHENG X, SCHUBERT I, JENUWEIN T AND JACOBSEN SE. 2004. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112: 308-315.

JACOB Y, FENG S, LEBLANC CA, BERNATAVICHUTE YV, STROUD H, COKUS S, JOHNSON LM, PELLEGRINI M, JACOBSEN SE AND MICHAELS SD. 2009. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nature structural & molecular biology 16: 763-768.

JARVIS P, BELZILE F AND DEAN C. 1997. Inefficient and incorrect processing of the Ac transposase transcript in iae1 and wild-type Arabidopsis thaliana. The Plant journal : for cell and molecular biology 11: 921-931.

JEFFERSON RA, KAVANAGH TA AND BEVAN MW. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO journal 6: 3901-3907.

JEGGO PA. 1998. DNA breakage and repair. Adv Genet 38: 185-218.

JEONG JH, SONG HR, KO JH, JEONG YM, KWON YE, SEOL JH, AMASINO RM, NOH B AND NOH YS. 2009. Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis. PloS one 4: e8033.

JIA Y, LISCH DR, OHTSU K, SCANLON MJ, NETTLETON D AND SCHNABLE PS. 2009. Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS genetics 5: e1000737.

JIANG D, YANG W, HE Y AND AMASINO RM. 2007. Arabidopsis relatives of the human lysine-specific Demethylase1 repress the expression of FWA and FLOWERING LOCUS C and thus promote the floral transition. The Plant cell 19: 2975-2987.

JIANG F, YANG M, GUO W, WANG X AND KANG L. 2012. Large-scale transcriptome analysis of retroelements in the migratory locust, Locusta migratoria. PloS one 7: e40532.

JIN J, LIU J, WANG H, WONG L AND CHUA NH. 2013. PLncDB: plant long non-coding RNA database. Bioinformatics 29: 1068-1071.

JOHNSON L, CAO X AND JACOBSEN S. 2002. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Current biology : CB 12: 1360-1367.

JONES MA, COVINGTON MF, DITACCHIO L, VOLLMERS C, PANDA S AND HARMER SL. 2010. Jumonji domain protein JMJD5 functions in both the plant and human circadian systems. Proceedings of the National Academy of Sciences of the United States of America 107: 21623-21628.

JORDAN IK, ROGOZIN IB, GLAZKO GV AND KOONIN EV. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19: 68-72.

JOUANNET V, MORENO AB, ELMAYAN T, VAUCHERET H, CRESPI MD AND MAIZEL A. 2012. Cytoplasmic Arabidopsis AGO7 accumulates in membrane-associated siRNA bodies and is required for ta-siRNA biogenesis. The EMBO journal 31: 1704-1713.

JUNG JH, LEE S, YUN J, LEE M AND PARK CM. 2014. The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. Plant science : an international journal of experimental plant biology 215-216: 29-38.

JURICA MS AND STODDARD BL. 1999. Homing endonucleases: structure, function and evolution. Cell Mol Life Sci 55: 1304-1326.

KALMYKOVA AI, KLENOV MS AND GVOZDEV VA. 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic acids research 33: 2052-2059.

KANKEL MW, RAMSEY DE, STOKES TL, FLOWERS SK, HAAG JR, JEDDELOH JA, RIDDLE NC, VERBSKY ML AND RICHARDS EJ. 2003. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics 163: 1109-1122.

KAPITONOV VV AND JURKA J. 2001. Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 98: 8714-8719.

KATO H, GOTO DB, MARTIENSSEN RA, URANO T, FURUKAWA K AND MURAKAMI Y. 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science 309: 467-469.

KATO M, MIURA A, BENDER J, JACOBSEN SE AND KAKUTANI T. 2003. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Current biology : CB 13: 421-426.

KIM CM PH, PARK SJ, CHON NS, JE BI, SUN B, PARK SH, PARK JY, LEE EJ, KIM MJ, CHUNG WS, LEE KH, LEE YS, LEE JJ, WON YJ, YI G, NAM MH, CHA YS, YUN DW, EUN MY, HAN CD. 2004. Rapid, large-scale generation of Ds transposant lines and analysis of the Ds insertion sites in rice. Plant J 39: 252-263.

KIM ED AND SUNG S. 2012. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends in plant science 17: 16-21.

KIM MK, SHIN JM, EUN HC AND CHUNG JH. 2009. The role of p300 histone acetyltransferase in UV-induced histone modifications and MMP-1 gene transcription. PloS one 4: e4864.

KIM YG, CHA J AND CHANDRASEGARAN S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America 93: 1156-1160.

KOBAYASHI S, GOTO-YAMAMOTO N AND HIROCHIKA H. 2004. Retrotransposon-induced mutations in grape skin color. Science 304: 982.

KONCZ C AND SCHELL J. 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. Mol Gen Genet 204: 383-396.

KONDO Y AND ISSA JP. 2003. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. The Journal of biological chemistry 278: 27658-27662.

KONG L, ZHANG Y, YE ZQ, LIU XQ, ZHAO SQ, WEI L AND GAO G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector

machine. Nucleic acids research 35: W345-349.

KOPREK T, MCELROY D, LOUWERSE J, WILLIAMS-CARRIER R AND LEMAUX PG. 2000. An efficient method for dispersing Ds elements in the barley genome as a tool for determining gene function. The Plant journal : for cell and molecular biology 24: 253-263.

KUNZE R, KUHN S, JONES JDG AND SCOFIELD SR. 1995. Somatic and Germinal Activities of Maize Activator (Ac) Transposase Mutants in Transgenic Tobacco. Plant Journal 8: 45-54.

KUNZE R AND STOCHAJ U LJ, STARLINGER P. 1987. Transcription of transposable element Activator (Ac) of Zea mays L. EMBO J 6: 1555-1563.

KUNZE R AND WEIL CF. 2002. The hAT and CACTA superfamilies of plant transposons. Mobile DNA II: 565-610.

LA H ET AL. 2011. A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. Proceedings of the National Academy of Sciences of the United States of America 108: 15498-15503.

LAMESCH P ET AL. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research 40: D1202-1210.

LANDER ES ET AL. 2001. Initial sequencing and analysis of the human genome. Nature 409: 860-921.

LANGMEAD B AND SALZBERG SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.

LANGMEAD B, TRAPNELL C, POP M AND SALZBERG SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.

LANS H, MARTEIJN JA AND VERMEULEN W. 2012. ATP-dependent chromatin remodeling in the DNA-damage response. Epigenetics Chromatin 5: 4.

LAURIA M AND ROSSI V. 2011. Epigenetic control of gene regulation in plants. Biochimica et biophysica acta 1809: 369-378.

LAW JA AND JACOBSEN SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics 11: 204-220.

LEE CG, DA SILVA CA, LEE JY, HARTL D AND ELIAS JA. 2008. Chitin regulation of immune responses: an old molecule with new roles. Current opinion in immunology 20: 684-689.

LEE JH AND PAULL TT. 2005. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science 308: 551-554.

LEE TF, GURAZADA SG, ZHAI J, LI S, SIMON SA, MATZKE MA, CHEN X AND MEYERS BC. 2012. RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. Epigenetics : official journal of the DNA Methylation Society 7: 781-795.

LI C ET AL. 2009a. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137: 509-521.

LI MZ AND ELLEDGE SJ. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4: 251-256.

LI R, LI Y, KRISTIANSEN K AND WANG J. 2008. SOAP: short oligonucleotide alignment program. Bioinformatics 24: 713-714.

LI R, YU C, LI Y, LAM TW, YIU SM, KRISTIANSEN K AND WANG J. 2009b. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967.

LI X ET AL. 2014. Histone demethylase KDM5B is a key regulator of genome stability. Proceedings of the National Academy of Sciences of the United States of America 111: 7096-7101.

LIBAULT M, WAN J, CZECHOWSKI T, UDVARDI M AND STACEY G. 2007. Identification of 118 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. Molecular plant-microbe interactions : MPMI 20: 900-911.

LICHTENSTEIN AV, MOISEEV VL AND ZABOIKIN MM. 1990. A procedure for DNA and RNA transfer to membrane filters avoiding weight-induced gel flattening. Analytical biochemistry 191: 187-191.

LIEBER MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry 79: 181-211.

LIPPMAN Z, MAY B, YORDAN C, SINGER T AND MARTIENSSEN R. 2003. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS biology 1: E67.

LISCH D. 2013. How important are transposons for plant evolution? Nature reviews Genetics 14: 49-61.

LISCH D, CAREY CC, DORWEILER JE AND CHANDLER VL. 2002. A mutation that prevents paramutation in maize also reverses Mutator transposon methylation and silencing. Proceedings of the National Academy of Sciences of the United States of America 99: 6130-6135.

LISSON R, HELLERT J, RINGLEB M, MACHENS F, KRAUS J AND HEHL R. 2010. Alternative splicing of the maize Ac transposase transcript in transgenic sugar beet (Beta vulgaris L.). Plant molecular biology 74: 19-32.

LISTER R, O'MALLEY RC, TONTI-FILIPPINI J, GREGORY BD, BERRY CC, MILLAR AH AND ECKER JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523-536.

LIU C, LU F, CUI X AND CAO X. 2010a. Histone methylation in higher plants. Annual review of plant biology 61: 395-420.

LIU F, MARQUARDT S, LISTER C, SWIEZEWSKI S AND DEAN C. 2010b. Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science 327: 94-97.

LIU F, QUESADA V, CREVILLEN P, BAURLE I, SWIEZEWSKI S AND DEAN C. 2007. The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Molecular cell 28: 398-407.

LIU J, JUNG C, XU J, WANG H, DENG S, BERNAD L, ARENAS-HUERTERO C AND CHUA NH. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant cell 24: 4333-4345.

LIU X, YU CW, DUAN J, LUO M, WANG K, TIAN G, CUI Y AND WU K. 2012. HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable elements silencing in Arabidopsis. Plant physiology 158: 119-129.

LLOYD A, PLAISIER CL, CARROLL D AND DREWS GN. 2005. Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proceedings of the National Academy of Sciences

of the United States of America 102: 2232-2237.

LOHSE M, BOLGER AM, NAGEL A, FERNIE AR, LUNN JE, STITT M AND USADEL B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic acids research 40: W622-627.

LU F, CUI X, ZHANG S, JENUWEIN T AND CAO X. 2011. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nature genetics 43: 715-719.

LU F, CUI X, ZHANG S, LIU C AND CAO X. 2010. JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. Cell research 20: 387-390.

LU F, LI G, CUI X, LIU C, WANG XJ AND CAO X. 2008. Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. J Integr Plant Biol 50: 886-896.

LU SX, KNOWLES SM, WEBB CJ, CELAYA RB, CHA C, SIU JP AND TOBIN EM. 2011b. The Jumonji C domain-containing protein JMJ30 regulates period length in the Arabidopsis circadian clock. Plant physiology 155: 906-915.

LUO M, LIU X, SINGH P, CUI Y, ZIMMERLI L AND WU K. 2012. Chromatin modifications and remodeling in plant abiotic stress responses. Biochimica et biophysica acta 1819: 129-136.

MA H, RAO L, WANG HL, MAO ZW, LEI RH, YANG ZY, QING H AND DENG YL. 2013a. Transcriptome analysis of glioma cells for the dynamic response to gamma-irradiation and dual regulation of apoptosis genes: a new insight into radiotherapy for glioblastomas. Cell Death Dis 4: e895.

MA X, LV S, ZHANG C AND YANG C. 2013b. Histone deacetylases and their functions in plants. Plant cell reports 32: 465-478.

MAKAREVICH G, LEROY O, AKINCI U, SCHUBERT D, CLARENZ O, GOODRICH J, GROSSNIKLAUS U AND KOHLER C. 2006. Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO reports 7: 947-952.

MANNUSS A, TRAPP O AND PUCHTA H. 2012. Gene regulation in response to DNA damage. Biochimica et biophysica acta 1819: 154-165.

MASCHERETTI I, BATTAGLIA R, MAINIERI D, ALTANA A, LAURIA M AND ROSSI V. 2013. The WD40-repeat proteins NFC101 and NFC102 regulate different aspects of maize development through chromatin modification. The Plant cell 25: 404-420.

MATZKE MA AND BIRCHLER JA. 2005. RNAi-mediated pathways in the nucleus. Nature reviews Genetics 6: 24-35.

MCCLINTOCK B. 1948. Mutable Loci in Maize. Carnegie Inst Wash Year Book 47: 155-169.

MCCLINTOCK B. 1949. Mutable loci in maize. Carnegie Inst Wash Year Book 48: 142-154.

MCCLINTOCK B. 1950. The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences of the United States of America 36: 344-355.

MCCLINTOCK B. 1951. Mutable loci in maize. Carnegie Inst Wash Year Book 50: 174-181.

MCCLINTOCK B. 1984. The significance of responses of the genome to challenge. Science 226: 792-801.

MEEK DW. 2009. Tumour suppression by p53: a role for the DNA damage response? Nat Rev Cancer 9: 714-723.

MEISTER G. 2013. Argonaute proteins: functional insights and emerging roles. Nature reviews Genetics 14: 447-459.

MENONI H, GASPARUTTO D, HAMICHE A, CADET J, DIMITROV S, BOUVET P AND ANGELOV D. 2007. ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes. Molecular and cellular biology 27: 5949-5956.

MITRA S, IZUMI T, BOLDOGH I, BHAKAT KK, HILL JW AND HAZRA TK. 2002. Choreography of oxidative damage repair in mammalian genomes. Free radical biology & medicine 33: 15-28.

MIURA A, YONEBAYASHI S, WATANABE K, TOYAMA T, SHIMADA H AND KAKUTANI T. 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411: 212-214.

MURASHIGE T AND SKOOG F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum 15: 473-497.

MURRAY MG AND THOMPSON WF. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic acids research 8: 4321-4325.

NAGATA T, YAMADA H, DU Z, TODORIKI S AND KIKUCHI S. 2005. Microarray analysis of genes that respond to gamma-irradiation in Arabidopsis. J Agric Food Chem 53: 1022-1030.

NENE V ET AL. 2007. Genome sequence of Aedes aegypti, a major arbovirus vector. Science 316: 1718-1723.

NG DW, WANG T, CHANDRASEKHARAN MB, ARAMAYO R, KERTBUNDIT S AND HALL TC. 2007. Plant SET domain-containing proteins: structure, function and regulation. Biochimica et biophysica acta 1769: 316-329.

NICOL JW, HELT GA, BLANCHARD SG, JR., RAJA A AND LORAINE AE. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25: 2730-2731.

OGIWARA H, UI A, OTSUKA A, SATOH H, YOKOMI I, NAKAJIMA S, YASUI A, YOKOTA J AND KOHNO T. 2011. Histone acetylation by CBP and p300 at double-strand break sites facilitates SWI/SNF chromatin remodeling and the recruitment of non-homologous end joining factors. Oncogene 30: 2135-2146.

ORGEL LE AND CRICK FH. 1980. Selfish DNA: the ultimate parasite. Nature 284: 604-607.

OSANAI-FUTAHASHI M, SUETSUGU Y, MITA K AND FUJIWARA H. 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, Bombyx mori. Insect Biochem Mol Biol 38: 1046-1057.

PANDEY R, MULLER A, NAPOLI CA, SELINGER DA, PIKAARD CS, RICHARDS EJ, BENDER J, MOUNT DW AND JORGENSEN RA. 2002. Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic acids research 30: 5036-5055.

PANDITA TK AND RICHARDSON C. 2009. Chromatin remodeling finds its place in the DNA double-strand break response. Nucleic acids research 37: 1363-1377.

PARK YJ, DIXIT A, YOO JW AND BENNETZEN J. 2004. Further evidence of microcolinearity between barley and rice genomes at two orthologous regions. Molecules and cells 17: 492-502.

PAULL TT AND LEE JH. 2005. The Mre11/Rad50/Nbs1 complex and its role as a DNA

double-strand break sensor for ATM. Cell Cycle 4: 737-740.

PECINKA A, DINH HQ, BAUBEC T, ROSA M, LETTNER N AND MITTELSTEN SCHEID O. 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. The Plant cell 22: 3118-3129.

PENA PV ET AL. 2008. Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. Journal of molecular biology 380: 303-312.

PETERSON PA. 1953. A mutable pale green locus in maize. Genetics 38: 682-683.

PONTES O, LI CF, COSTA NUNES P, HAAG J, REAM T, VITINS A, JACOBSEN SE AND PIKAARD CS. 2006. The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell 126: 79-92.

PUCHTA H. 1999a. Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. Genetics 152: 1173-1181.

PUCHTA H. 1999b. Use of I-Sce I to induce DNA double-strand breaks in Nicotiana. Methods Mol Biol 113: 447-451.

QI Y, HE X, WANG XJ, KOHANY O, JURKA J AND HANNON GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443: 1008-1012.

QU F, YE X AND MORRIS TJ. 2008. Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. Proceedings of the National Academy of Sciences of the United States of America 105: 14732-14737.

QUESTA JI, WALBOT V AND CASATI P. 2010. Mutator transposon activation after UV-B involves chromatin remodeling. Epigenetics : official journal of the DNA Methylation Society 5: 352-363.

REAM TS, HAAG JR, WIERZBICKI AT, NICORA CD, NORBECK AD, ZHU JK, HAGEN G, GUILFOYLE TJ, PASA-TOLIC L AND PIKAARD CS. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Molecular cell 33: 192-203.

REBOLLAR E, VALADEZ-GRAHAM V, VAZQUEZ M, REYNAUD E AND ZURITA M. 2006. Role of the p53 homologue from Drosophila melanogaster in the maintenance of histone H3 acetylation and response to UV-light irradiation. FEBS letters 580: 642-648.

REIDT W, WURZ R, WANIECK K, CHU HH AND PUCHTA H. 2006. A homologue of the breast cancer-associated gene BARD1 is involved in DNA repair in plants. The EMBO journal 25: 4326-4337.

RICAUD L, PROUX C, RENOU JP, PICHON O, FOCHESATO S, ORTET P AND MONTANE MH. 2007. ATM-mediated transcriptional and developmental responses to gamma-rays in Arabidopsis. PloS one 2: e430.

RICHARDS S ET AL. 2008. The genome of the model beetle and pest Tribolium castaneum. Nature 452: 949-955.

ROBINSON MD, MCCARTHY DJ AND SMYTH GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

ROCHAIX JD, RAHIRE M AND MICHEL F. 1985. The chloroplast ribosomal intron of Chlamydomonas reinhardii codes for a polypeptide related to mitochondrial maturases.

Nucleic acids research 13: 975-984.

ROMANISH MT, LOCK WM, VAN DE LAGEMAAT LN, DUNN CA AND MAGER DL. 2007. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS genetics 3: e10.

ROS F AND KUNZE R. 2001. Regulation of activator/dissociation transposition by replication and DNA methylation. Genetics 157: 1723-1733.

ROTHKAMM K, KRUGER I, THOMPSON LH AND LOBRICH M. 2003. Pathways of DNA double-strand break repair during the mammalian cell cycle. Molecular and cellular biology 23: 5706-5715.

ROWLEY MJ, AVRUTSKY MI, SIFUENTES CJ, PEREIRA L AND WIERZBICKI AT. 2011. Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS genetics 7: e1002120.

RUDIN CM AND THOMPSON CB. 2001. Transcriptional activation of short interspersed elements by DNA-damaging agents. Genes Chromosomes Cancer 30: 64-71.

SACERDOT C, MERCIER G, TODESCHINI AL, DUTREIX M, SPRINGER M AND LESAGE P. 2005. Impact of ionizing radiation on the life cycle of Saccharomyces cerevisiae Ty1 retrotransposon. Yeast 22: 441-455.

SAITO Y, SUZUKI H, TSUGAWA H, NAKAGAWA I, MATSUZAKI J, KANAI Y AND HIBI T. 2009. Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. Oncogene 28: 2738-2744.

SAMBROOK JF AND RUSSELL DW. 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press.

SARGENT RG, BRENNEMAN MA AND WILSON JH. 1997. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. Molecular and cellular biology 17: 267-277.

SASAKI T, KOBAYASHI A, SAZE H AND KAKUTANI T. 2012. RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1. The Plant journal : for cell and molecular biology 70: 750-758.

SAZE H AND KAKUTANI T. 2011. Differentiation of epigenetic modifications between transposons and genes. Current opinion in plant biology 14: 81-87.

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.

SCHLEGEL BP, JODELKA FM AND NUNEZ R. 2006. BRCA1 promotes induction of ssDNA by ionizing radiation. Cancer research 66: 5181-5189.

SCHMIDT R AND WILLMITZER L. 1989. The Maize Autonomous Element Activator (Ac) Shows a Minimal Germinal Excision Frequency of 0.2-Percent-0.5-Percent in Transgenic Arabidopsis-Thaliana Plants. Molecular & General Genetics 220: 17-24.

SCHMITZ RJ, SCHULTZ MD, LEWSEY MG, O'MALLEY RC, URICH MA, LIBIGER O, SCHORK NJ AND ECKER JR. 2011. Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373.

SCHNABLE PS ET AL. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

SCHWARTZ D AND DENNIS E. 1986. Transposase Activity of the Ac Controlling Element in

Maize Is Regulated by Its Degree of Methylation. Molecular & General Genetics 205: 476-482.

SCOFIELD SR, ENGLISH JJ AND JONES JD. 1993. High level expression of the Activator transposase gene inhibits the excision of Dissociation in tobacco cotyledons. Cell 75: 507-517.

SELA N, KIM E AND AST G. 2010. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. Genome biology 11: R59.

SEO JK, WU J, LII Y, LI Y AND JIN H. 2013. Contribution of small RNA pathway components in plant immunity. Molecular plant-microbe interactions : MPMI 26: 617-625.

SHAKED H, AVIVI-RAGOLSKY N AND LEVY AA. 2006. Involvement of the Arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. Genetics 173: 985-994.

SHI Q, PARKS AR, POTTER BD, SAFIR IJ, LUO Y, FORSTER BM AND PETERS JE. 2008. DNA damage differentially activates regional chromosomal loci for Tn7 transposition in Escherichia coli. Genetics 179: 1237-1250.

SHI Y, LAN F, MATSON C, MULLIGAN P, WHETSTINE JR, COLE PA AND CASERO RA. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119: 941-953.

SHRIVASTAV M, DE HARO LP AND NICKOLOFF JA. 2008. Regulation of DNA double-strand break repair pathway choice. Cell research 18: 134-147.

SHUKLA V ET AL. 2010. BRCA1 affects global DNA methylation through regulation of DNMT1. Cell research 20: 1201-1215.

SIJEN T AND PLASTERK RH. 2003. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature 426: 310-314.

SIOMI MC, SATO K, PEZIC D AND ARAVIN AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology 12: 246-258.

SLOTKIN RK AND MARTIENSSEN R. 2007. Transposable elements and the epigenetic regulation of the genome. Nature reviews Genetics 8: 272-285.

SPANUDAKIS E AND JACKSON S. 2014. The role of microRNAs in the control of flowering time. Journal of experimental botany 65: 365-380.

SRIDHAR VV, KAPOOR A, ZHANG K, ZHU J, ZHOU T, HASEGAWA PM, BRESSAN RA AND ZHU JK. 2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature 447: 735-738.

STROUD H, HALE CJ, FENG S, CARO E, JACOB Y, MICHAELS SD AND JACOBSEN SE. 2012. DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis. PLoS genetics 8: e1002808.

SUGIMOTO K, TAKEDA S AND HIROCHIKA H. 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. The Plant cell 12: 2511-2528.

SUN Y, JIANG X, XU Y, AYRAPETOV MK, MOREAU LA, WHETSTINE JR AND PRICE BD. 2009. Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. Nature cell biology 11: 1376-1382.

SYMINGTON LS AND GAUTIER J. 2011. Double-strand break end resection and repair pathway choice. Annual review of genetics 45: 247-271.

SZCZEPEK M, BRONDANI V, BUCHEL J, SERRANO L, SEGAL DJ AND CATHOMEN T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25: 786-793.

TAKEDA S, SUGIMOTO K, OTSUKI H AND HIROCHIKA H. 1998. Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate. Plant molecular biology 36: 365-376.

TAKEDA S, SUGIMOTO K, OTSUKI H AND HIROCHIKA H. 1999. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. The Plant journal : for cell and molecular biology 18: 383-393.

TAMURA K, ADACHI Y, CHIBA K, OGUCHI K AND TAKAHASHI H. 2002. Identification of Ku70 and Ku80 homologues in Arabidopsis thaliana: evidence for a role in the repair of DNA double-strand breaks. The Plant journal : for cell and molecular biology 29: 771-781.

THIERRY A AND DUJON B. 1992. Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. Nucleic acids research 20: 5625-5631.

THIMM O, BLASING O, GIBON Y, NAGEL A, MEYER S, KRUGER P, SELBIG J, MULLER LA, RHEE SY AND STITT M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant journal : for cell and molecular biology 37: 914-939.

THORSTENSEN T, GRINI PE AND AALEN RB. 2011. SET domain proteins in plant development. Biochimica et biophysica acta 1809: 407-420.

THORSTENSEN T, GRINI PE, MERCY IS, ALM V, ERDAL S, AASLAND R AND AALEN RB. 2008. The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). Plant molecular biology 66: 47-59.

THYKJAER T, STILLER J, HANDBERG K, JONES J AND STOUGAARD J. 1995. The maize transposable element Ac is mobile in the legume Lotus japonicus. Plant molecular biology 27: 981-993.

TOVKACH A, ZEEVI V AND TZFIRA T. 2009. A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. The Plant journal : for cell and molecular biology 57: 747-757.

TRAN RK ET AL. 2005. Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis. Genome biology 6: R90.

TRAPNELL C, ROBERTS A, GOFF L, PERTEA G, KIM D, KELLEY DR, PIMENTEL H, SALZBERG SL, RINN JL AND PACHTER L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562-578.

TREWICK SC, MCLAUGHLIN PJ AND ALLSHIRE RC. 2005. Methylation: lost in hydroxylation? EMBO reports 6: 315-320.

TSUKADA Y, FANG J, ERDJUMENT-BROMAGE H, WARREN ME, BORCHERS CH, TEMPST P AND ZHANG Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. Nature 439: 811-816.

TSUKAHARA S, KOBAYASHI A, KAWABE A, MATHIEU O, MIURA A AND KAKUTANI T. 2009.

Bursts of retrotransposition reproduced in Arabidopsis. Nature 461: 423-426.
TURCOTTE K, SRINIVASAN S AND BUREAU T. 2001. Survey of transposable elements from rice genomic sequences. Plant Journal 25: 169-179.
TUTEJA N, AHMAD P, PANDA BB AND TUTEJA R. 2009. Genotoxic stress in plants: shedding light on DNA damage, repair and DNA repair helicases. Mutation research 681: 134-149.
UZIEL T, LERENTHAL Y, MOYAL L, ANDEGEKO Y, MITTELMAN L AND SHILOH Y. 2003. Requirement of the MRN complex for ATM activation by DNA damage. The EMBO journal 22: 5612-5621.
VAGIN VV, KLENOV MS, KALMYKOVA AI, STOLYARENKO AD, KOTELNIKOV RN AND GVOZDEV VA. 2004. The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. RNA biology 1: 54-58.
VAN ATTIKUM H, BUNDOCK P, OVERMEER RM, LEE LY, GELVIN SB AND HOOYKAAS PJ. 2003. The Arabidopsis AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. Nucleic acids research 31: 4247-4255.
VAN SLUYS MA, TEMPE J AND FEDOROFF N. 1987. Studies on the introduction and mobility of the maize Activator element in Arabidopsis thaliana and Daucus carota. The EMBO journal 6: 3881-3889.
VANYUSHIN BF AND ASHAPKIN VV. 2011. DNA methylation in higher plants: past, present and future. Biochimica et biophysica acta 1809: 360-368.
VAUCHERET H. 2008. Plant ARGONAUTES. Trends in plant science 13: 350-358.
WALKER EL, ROBBINS TP, BUREAU TE, KERMICLE J AND DELLAPORTA SL. 1995. Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. The EMBO journal 14: 2350-2363.
WAN G, HU X, LIU Y, HAN C, SOOD AK, CALIN GA, ZHANG X AND LU X. 2013. A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. The EMBO journal.
WANG L ET AL. 2014. Genome-wide screening and identification of long noncoding RNAs and their interaction with protein coding RNAs in bladder urothelial cell carcinoma. Cancer letters 349: 77-86.
WANG L, HEINLEIN M AND KUNZE R. 1996. Methylation pattern of Activator transposase binding sites in maize endosperm. The Plant cell 8: 747-758.
WANG L, MAI YX, ZHANG YC, LUO Q AND YANG HQ. 2010. MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in Arabidopsis. Molecular plant 3: 794-806.
WANG X, WEIGEL D AND SMITH LM. 2013. Transposon variants and their effects on gene expression in Arabidopsis. PLoS genetics 9: 7.
WANG Z, GERSTEIN M AND SNYDER M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics 10: 57-63.
WATERSTON RH ET AL. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.
WEI W, BA Z, GAO M, WU Y, MA Y, AMIARD S, WHITE CI, RENDTLEW DANIELSEN JM, YANG YG AND QI Y. 2012. A role for small RNAs in DNA double-strand break repair. Cell 149: 101-112.

WEIGEL D AND GLAZEBROOK J. 2002. Arabidopsis: A Laboratory Manual.

WEIL CF AND KUNZE R. 2000. Transposition of maize Ac/Ds transposable elements in the yeast Saccharomyces cerevisiae. Nature genetics 26: 187-190.

WEST CE, WATERWORTH WM, JIANG Q AND BRAY CM. 2000. Arabidopsis DNA ligase IV is induced by gamma-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. Plant Journal 24: 67-78.

WIERZBICKI AT. 2012. The role of long non-coding RNA in transcriptional gene silencing. Current opinion in plant biology 15: 517-522.

WIERZBICKI AT, HAAG JR AND PIKAARD CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635-648.

WIERZBICKI AT, REAM TS, HAAG JR AND PIKAARD CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature genetics 41: 630-634.

WU J, LU LY AND YU X. 2010. The role of BRCA1 in DNA damage response. Protein Cell 1: 117-123.

XIE A ET AL. 2007. Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. Molecular cell 28: 1045-1057.

XIE H, WANG M, BONALDO MDE F, SMITH C, RAJARAM V, GOLDMAN S, TOMITA T AND SOARES MB. 2009. High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. Nucleic acids research 37: 4331-4340.

XIN M, WANG Y, YAO Y, SONG N, HU Z, QIN D, XIE C, PENG H, NI Z AND SUN Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biol 11: 61.

YODER JI, PALYS J, ALPERT K AND LASSNER M. 1988. Ac Transposition in Transgenic Tomato Plants. Molecular & General Genetics 213: 291-296.

YOSHIDA K AND MIKI Y. 2004. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci 95: 866-871.

YOSHIYAMA K, CONKLIN PA, HUEFNER ND AND BRITT AB. 2009. Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. Proceedings of the National Academy of Sciences of the United States of America 106: 12843-12848.

YU X, LI L, GUO M, CHORY J AND YIN Y. 2008. Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 105: 7618-7623.

YUN MH AND HIOM K. 2009. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. Nature 459: 460-463.

ZEMACH A, MCDANIEL IE, SILVA P AND ZILBERMAN D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916-919.

ZHANG J, ZHANG S, HAN S, WU T, LI X, LI W AND QI L. 2012. Genome-wide identification of microRNAs in larch and stage-specific modulation of 11 conserved microRNAs and their targets during somatic embryogenesis. Planta 236: 647-657.

ZHANG K, SRIDHAR VV, ZHU J, KAPOOR A AND ZHU JK. 2007. Distinctive core histone

post-translational modification patterns in Arabidopsis thaliana. PloS one 2: e1210.
ZHANG YC AND CHEN YQ. 2013. Long noncoding RNAs: new regulators in plant development. Biochemical and biophysical research communications 436: 111-114.
ZHAO J, SUN BK, ERWIN JA, SONG JJ AND LEE JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322: 750-756.
ZHAO Y, HULL AK, GUPTA NR, GOSS KA, ALONSO J, ECKER JR, NORMANLY J, CHORY J AND CELENZA JL. 2002. Trp-dependent auxin biosynthesis in Arabidopsis: involvement of cytochrome P450s CYP79B2 and CYP79B3. Genes & development 16: 3100-3112.
ZHAO Z, YU Y, MEYER D, WU C AND SHEN WH. 2005. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nature cell biology 7: 1256-1260.
ZHU JK. 2009. Active DNA demethylation mediated by DNA glycosylases. Annual review of genetics 43: 143-166.
ZHUANG J ET AL. 2010. DNA demethylation in retinal neurocytes contributes to the upregulation of DNA repair protein, Ku80. Neuroreport 21: 282-286.
ZIERHUT C AND DIFFLEY JF. 2008. Break dosage, cell cycle stage and DNA replication influence DNA double strand break response. The EMBO journal 27: 1875-1885.
ZILBERMAN D, CAO X AND JACOBSEN SE. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299: 716-719.
ZUO J, NIU QW AND CHUA NH. 2000. Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant Journal 24: 265-273.

8 Appendix

Table S1: Primers for generating constructs of estradiol induced cleavage system

Primer names	Sequences (5' → 3')	Comments
ZW_P11	TCGACTCTAGCCTCGAGGCCACCATGAAGAACATTAAGAA	SUNGENE forw.
ZW_P12	GAGGCCTGGATCGACTAGTTCATGCTCGAGTCACTTCAGG	SUNGENE rev.
ZW_P13	TAAGCTTGCATGCCTGCAGGAT	G10-90 forw.
ZW_P14	TTCATGGATCCCAGCGTGTCCTCT	G10-90 rev.
ZW_P15	AGGACACGCTGGGATCCATGAAAAACATCAAAAAAAACCA	I-Sce I forw.
ZW_P16	ATGCCATAATACTCGAACATTATTTCAGGAAAGTTTCGGA	I-Sce I rev.
ZW_P17	GAAACTTTCCTGAAATAATGTTCGAGTATTATGGCATTGG	Hyg forw.
ZW_P18	GCCGAATTGAATTCGAGCTCCTGTCGAGGGGGGATCAAT	Hyg rev.
ZW_P19	CCTGCAGGATATCGTGGATCCATTTGGGGCCCACATAC	2'-1'nptII. forw.
ZW_P20	CATGGTTATGAAATTCAGATGCTAGTGT	2'-1'nptII. rev.
ZW_P21	AGCATCTGAATTTCATAACCATGTTCTTCCCCTCCTGA	QQR-Target forw.
ZW_P22	AGCATCTGAATTTCATAACCATGAGCACCAACTCCTGA	ZFN3-Target forw.
ZW_P23	TAGGGATAACAGGGTAATTTACGTCCTGTAGAAACCCC	I SceI-Target forw.1
ZW_P24	AGCATCTGAATTTCATAACCATGTAGGGATAACAGGGTAA	I SceI-Target forw.2
ZW_P25	GGGAGGCCTGGATCGACTAGTCATTGTTTGCCTCCCTG	GUS rev.
ZW_P36	AAACTGAAGGCGGGAAAC	pMDC7 seq 1
ZW_P37	TGTGTGGGCAATGAAACTG	pMDC7 seq 2
ZW_P38	ATAGAAGGCGGCGGTGGA	nptII seq

Table S2: Primers for *Ac*

Primer names	Sequences (5' → 3')	Comments
ZW_P59	CTCACTTTCCAAT**C**AAGTCCCGTGTCACTGC	Ac-a1886c forw.
ZW_P60	GACACGGGACTT**G**ATTGGAAAGTGAGGGC	Ac-a1886c rev.
ZW_P61	CTCACTGAACTCCTATCTGG**A**ACTCAATATTCCACTGC	Ac-t2737a forw.
ZW_P62	GCAGTGGAATATTGAGT**T**CCAGATAGGAGTTCAGTGAG	Ac-t2737a rev.
ZW_P63	ATCATGGTGG**C**GGGGAAGGGTTGCAG	Ac-a3239c forw.
ZW_P64	ACCCTTCCCC**G**CCACCATGATAAAATATCAAACTG	Ac-a3239c rev.
ZW_P80	aacacgggggactcttgacGGCTACGACTCCATTCCT	Ac-mut-SLIC-1F
ZW_P81	GAAAGAAAGCACCATCACAA	Ac-mut-SLIC-1R
ZW_P82	TTGTGATGGTGCTTTCTTTC	Ac-mut-SLIC-2F
ZW_P83	CACTATGCTTCAAAAGGGGT	Ac-mut-SLIC-2R
ZW_P84	ACCCCTTTTGAAGCATAGTG	Ac-mut-SLIC-3F
ZW_P85	gggaaattcgagctggtcacCTATTACAACCAAGGCTCAT	Ac-mut-SLIC-3R
ZW_P144	ttgaccatggtagatctGGCTACGACTCCATTCCT	Ac-mut-1F2
ZW_P166	TCTCAATGTTCCACTTAACC	Ac-M3-RL-F1R (NarI)
ZW_P167	TTGGATTGATGATGATTGGTGTC	Ac-M3-RL-F2F (NarI)
ZW_P168	ATGAACTTTGTATGAATCACCA	Ac-M3-RL-F2R (SphI)

ZW_P169	GCAATGAGTGAAAAGTTTGAGAAAT	Ac-M3-RL-F3F (SphI)
ZW_P197	TTGCGTCTTCATCCTCAT	Ac transcripts-rev.
SF o520	GCCTCTTCTAGTCGGATT	Ac-probe-forw.
Sf o16	AAACTATGTGATGTTCTCAAGTGA	Ac-probe-rev.
SF o06L	GAGGTCGATGGAAAGAAATACGTTC	Ac transcript-forw.

Table S3: Primers for qRT-PCR

Primer names	Sequences (5' → 3')	Comments	Primer names	Sequences (5' → 3')	Comments
ZW_P150	TGTCGATGGGAAATCATGCAAACC	CMT3-Qpcr-F	**ZW_P307**	AGCGCTGAAGGAAGCTACAAGC	Mutator-like-qRT-F
ZW_P151	TTCTGCCCGTGTGACAACAGTG	CMT3-Qpcr-R	**ZW_P308**	ACACACCGTCCCAAGATTTGTTTC	Mutator-like-qRT-R
ZW_P152	AAAGGTGTCTACGTGGGCGTTG	DRD1-Qpcr-F	**ZW_P309**	ACTTCAATTCACTGCCCTCTTTCG	Spm-like-qRT-F
ZW_P153	AGTCTCCCTTGCTACATCCTTGG	DRD1-Qpcr-R	**ZW_P310**	AGGTCTCAGGTTGGATCTGACAGG	Spm-like-qRT-R
ZW_P154	TTGCTATATGGACGCCAGGTGAG	ROS1-Qpcr-F	**ZW_P311**	ACGAAGAGTTCAGCCCTGGTTC	gypsy-like-qRT-F
ZW_P155	TGCACGTACTAACAGACGGTTG	ROS1-Qpcr-R	**ZW_P312**	TTTCCAGCGTTCGGTTTGGC	gypsy-like-qRT-R
ZW_P156	ACAGTCTGAGGGAAGCAAACTCAC	DME-Qpcr-F	**ZW_P313**	GGACTTGAAGGTGTTGGTTGAAGC	Spm-like (CAC1) -qRT-F
ZW_P157	TGGCAGTCCGACAAGGTATCAG	DME-Qpcr-R	**ZW_P314**	GATGCAGGCCATGCAATCATCTG	Spm-like (CAC1) -qRT-R
ZW_P158	ATCTAGCTGGTGTAGCCGTGAC	DRM2-Qpcr-F	**ZW_P315**	TTTCTCCTGTTCCGGACGTTGC	ATMUDR-qRT-1F
ZW_P159	AACCTCGTCTGAGAAGCCCATC	DRM2-Qpcr-R	**ZW_P316**	GCTGCTGACTAGGAGGAAACAGTG	ATMUDR-qRT-1R
ZW_P160	TCTCGAAATGGGTTACCAGGTGAG	MET1-Qpcr-F	**ZW_P317**	TAACGCCGTTGGAAAGAGATGC	ATMUDR-qRT-2F
ZW_P161	CAAAGACATGCATCGGCTCAGG	MET1-Qpcr-R	**ZW_P318**	ACTGACCCATCAGACGTGTCTTC	ATMUDR-qRT-2R
ZW_P162	TGGCCAAGGGCAGTTTCATCAAG	DDM1-Qpcr-F	**ZW_P319**	AGATGGTGCGCCTATTCCTCTTC	CACTA-like-qRT-F
ZW_P163	AACGCCAGTATGTCCTCTTCCTC	DDM1-Qpcr-R	**ZW_P320**	TCCAGCTCTTTCACTCGGACAG	CACTA-like-qRT-F
ZW_P227	AGCAGTCGATTTATGGCGATGACC	Ku70-qRT-F	**ZW_P321**	ACCACTCCTAGCTTTGGTGATCTG	beta-6-tubulin-qRT-F
ZW_P228	CCACAGTCAAGTCCTTCAGCTTTC	Ku70-qRT-R	**ZW_P322**	AGGTTCACTGCGAGCTTCCTCA	beta-6-tubulin-qRT-R
ZW_P231	CAAGTCTCGTGGCTTTCAGGAG	ASH1-qRT-F	**ZW_P323**	TGAGCACGCTCTTCTTGCTTTCA	EF1a-qRT-F
ZW_P232	TCAGATGTAAGCTCGTCCTCTGC	ASH1-qRT-R	**ZW_P324**	GGTGGTGGCATCCATCTTGTTACA	EF1a-qRT-R
ZW_P233	ACGCTGCGGAAAGATTCGAGAC	LDL1-qRT-F	**ZW_P325**	GTTGCAGGTGTTGAAGCTAGAGGT	APT1-qRT-F
ZW_P234	TCCGCGTAGTATCTGCAAGACC	LDL1-qRT-R	**ZW_P326**	TGGCACCAATAGCCAACGCAATAG	APT1-qRT-R
ZW_P239	AGAGCAAGCGTTTGACAGGTAAG	EFS-qRT-F	**ZW_P327**	TACACGCGATAGAGAGCGCAAG	ATDMAP1-qRT-F
ZW_P240	ACAACCCAACCTACCATCAGGTG	EFS-qRT-R	**ZW_P328**	ACACATCAGGTTCTTCTGCACGAC	ATDMAP1-qRT-R
ZW_P241	CCGGAATACCGTGTTGCAGGTTAG	REF6-qRT-F	**ZW_P329**	GCCTTGAGGCTATCAGGTATTGGG	LDL3-qRT-F
ZW_P242	CCCGAATGATAAGCTCCCGGAAAG	REF6-qRT-R	**ZW_P330**	ACGAATATCACGGCTCGTGTGTG	LDL3-qRT-R
ZW_P243	TGCTGTCCTGTGTCTCAGGTTG	IBM1-qRT-F	**ZW_P331**	TGCGGAAGCTTACAGTTGGGATG	ATMetnase-qRT-F
ZW_P244	TGAAGGTCCAAGGTTCAATGCC	IBM1-qRT-R	**ZW_P332**	AGCTGCTTCAGAAACAACCTCTG	ATMetnase-qRT-R
ZW_P249	TCCTGCGATGCGTAGTGTTCTC	Ku80-qRT-F	**ZW_P333**	TCATCTGTGAATATGCAGGTGAGC	SUVR3-qRT-F
ZW_P250	AGCAAGCTCGTTTCCAGTTTCTTC	Ku80-qRT-R	**ZW_P334**	AAGCTTGTCCTGAAGGGAGGTG	SUVR3-qRT-R
ZW_P251	CCCGCTCTTACAGCTACAAGGTTC	RAD50-qRT-F	**ZW_P335**	AGTCAGGAAACAGAGCGCATGG	AtRAD21.1-qRT-F
ZW_P252	TTGACCTGCACTGCATCTTCCTC	RAD50-qRT-R	**ZW_P336**	GAAGGTCCTCTTCCATGTCAAACC	AtRAD21.1-qRT-R
ZW_P253	AGCAAGAGGCGGTAATACAGTTTG	SAM-qRT-F	**ZW_P337**	TGAAAGTGCGCCAACTGAAGAAG	ATRAD18-qRT-F
ZW_P254	CAAGAACAGCCTCTCCTTTCCG	SAM-qRT-R	**ZW_P338**	CGTGCATCTCAAGTTCCTTTGTCG	ATRAD18-qRT-R
ZW_P255	AGCCACATGTCTTCATGTTCTGC	SNM1-qRT-F	**ZW_P339**	GTGCCATTCAGATTGACCCAGAAC	ATRAD3-qRT-F
ZW_P256	TCTCGGAGTAAGTCCAACCTGTG	SNM1-qRT-R	**ZW_P340**	TGCCCTCATATCCAGTGATGCC	ATRAD3-qRT-R

ZW_P257	AACAAATCTCAGCCTCGGGTTAC	MRE11-qRT-F	**ZW_P467**	CCATCGAGCGGAAACCAAACAAG	AGO7-qRT_F
ZW_P258	AGAAGTTGTTCCGCTTGAGAGGTC	MRE11-qRT-R	**ZW_P468**	GCCGGTCAAGTTTAAGCTTTGGTG	AGO7-qRT_R
ZW_P259	TTGGCTTCAAGTGAGAACAGAGC	LIG4-qRT-F	**ZW_P469**	TGTCCACTCGCTCTCCTATGTG	AGO4-qRT_F
ZW_P260	TGACCCACTTCATCTCCTGAGC	LIG4-qRT-R	**ZW_P470**	CATAGCAGATCGGCGCAACAAC	AGO4-qRT_R
ZW_P261	GCTGCTAATCGTGAATGCGATCC	MEA-qRT-F	**ZW_P525**	GCTTGGCAGGCTACATCAACAC	TE_AT3G28160-qRT-F
ZW_P262	AGAGTGCCATCTCCACAGCTAAG	MEA-qRT-R	**ZW_P526**	AACCGTCGCTTGAGTCGCTATC	TE_AT3G28160-qRT-R
ZW_P263	ACTTGTCGAGTGGCATATCATCCG	ATX2-qRT-F	**ZW_P527**	TTACAGGCTGGTTTCGGTTGGG	TE_AT2G11240-qRT-F
ZW_P264	GCCTATCCTCATCTGCAAGCTC	ATX2-qRT-R	**ZW_P528**	TCATCGCTGCAGCTTTATGATGGG	TE_AT2G11240-qRT-R
ZW_P267	AGACGACGAGAAACCGCCTTTG	SUVH6-qRT-F	**ZW_P555**	ACAGCTTTCAGTCTGCATTGG	ATM-midQRT-F
ZW_P268	ACTTTGGAGGAACTGGTCTGCAC	SUVH6-qRT-R	**ZW_P556**	GCCCATGATAATGTGATACATCCT	ATM-midQRT-R
ZW_P269	TGTAGAAACTGCTGGGTCATTGG	CLF-qRT-F	**ZW_P557**	CGGTGCTCACAAAGAGCAACTATG	COPIA78qrt-F
ZW_P270	AATTATCGCCTCTTTGGCTTGGG	CLF-qRT-R	**ZW_P558**	ATCCTTGATAGATTAGACAGAGAGCT	COPIA78qrt-R
ZW_P271	ACCAACAGAACATGTAGACTTCCG	ATX1-qRT-F	**ZW_P561**	TGGGATGCCTACTCAGGAAAGAGG	WRR4-qRT-F
ZW_P272	AGAGTCTGTCACAACCTTTCCAAG	ATX1-qRT-R	**ZW_P562**	TCTTGAGATTGGCAAGGCACTGAG	WRR4-qRT-R
ZW_P273	TTCCTTCATTTCTGGCGTTGGG	SUVH5-qRT-F	**ZW_P563**	AACGATTGTTGGCTGCCGTTG	ATRE1-qRT-F
ZW_P274	ACCGCGCAAGGAACTAAGTATTG	SUVH5-qRT-R	**ZW_P564**	AAGACGCGCCTTGTATCGATTTAG	ATRE1-qRT-R
ZW_P275	TCTGCCTGGATGCCACAATCTG	SUVR4-qRT-F	**ZW_P577**	TTCCGCTCTGGAAAGACTCAGC	RAD51-qRT-F
ZW_P276	TGTTTGCATCCTCGCATCTGTG	SUVR4-qRT-R	**ZW_P578**	ACCTCCTTGATCCATGGGAAGTTG	RAD51-qRT-R
ZW_P277	ACCCAATTGCTACGCTAAGGTG	SWN-qRT-F	**ZW_P593**	TTAACCATACTCCCTATTTCAAGCA	AT1TE69975-qRT-F
ZW_P278	AAGCTCTTCGCTAGCTTCTATTCG	SWN-qRT-R	**ZW_P594**	TTGGGATATGTAGAACACTAGAAAACA	AT1TE69975-qRT-F
ZW_P279	TGCTTGGTGGGTTGCCAGATTG	ATXR3-qRT-F	**ZW_P599**	GCTAACTAAAATAGCGGCATCCTAC	AT5TE71740-qRT-R
ZW_P280	CTCGAAATTGATGAACCGGACCAG	ATXR3-qRT-R	**ZW_P600**	GACGAAATGTTGAGGGTTAACAAC	AT5TE71740-qRT-R
ZW_P281	TTGCACAGAGTGAGATTGAGGAAG	ASHR3-qRT-F	**ZW_P607**	ATTTTAACCCACTATTCACCACG	AT1TE69990-qRT-F
ZW_P282	AGGCGGATCTTCCTTCACAACAG	ASHR3-qRT-R	**ZW_P608**	CGGGTTAAATCTTCACTTTGCTA	AT1TE69990-qRT-R
ZW_P283	AGAGAAAGGTGGGATGCAGGTG	ATXR5-qRT-F	**ZW_P617**	ACGTAGCGATTGTTGTGGACAGTG	AT2G04460-qRT-F
ZW_P284	TACCACAACGAGAGGAGGGCATTC	ATXR5-qRT-R	**ZW_P618**	CTCACTCAACTATGAGCGTTCGTG	AT2G04460-qRT-R
ZW_P285	GCAAGGTCTTGGTGGAAGACAGAG	KYP-qRT-F	**ZW_P619**	TCCACCATCGTCTCCAACGCTAAC	AT2G14230-qRT-F
ZW_P286	TGCAGAACTCTGGCGCATTCTC	KYP-qRT-R	**ZW_P620**	GCACGACCAGGACTATCAAGAAGC	AT2G14230-qRT-R
ZW_P287	TTGCCCTGACAGACGCAGTAAC	ATXR6-qRT-F	**ZW_P629**	ACATGGTCACATCTTGTTGGACA	AT1TE04710-qRT-F
ZW_P288	TCTGCTTCTTCCTCCCTTCTGG	ATXR6-qRT-R	**ZW_P630**	ACATATTGTTCTTGGGGCTTGGA	AT1TE04710-qRT-R
ZW_P289	TCTGGGTCCACTCTTCAAGGTTAC	JMJ15-qRT-F	**ZW_P631**	CTCTACCCTTAACGGTTTGCACT	AT1TE04720-qRT-F
ZW_P290	TGGTTGCTGCGTCTCATGATCTC	JMJ15-qRT-R	**ZW_P632**	TGGAGAAAGATCATGGGTTGAGG	AT1TE04720-qRT-R
ZW_P293	TCTTGGTCCATTGTTCAGGGTCTC	JMJ14-qRT-F	**ZW_P657**	GCCAGGACCAAGTGGAATCA	JMJ30-qRT-F
ZW_P294	GAGCCTTTGTGTCACCATTTGCC	JMJ14-qRT-R	**ZW_P658**	ACAAAGGATGCTGGGCAAGA	JMJ30-qRT-R
ZW_P295	TGGCATTCCCTGCTGTAGGTTG	ELF6-qRT-F	**ZW_P660**	CATTCGTGATCTCTCCCACATCT	DRM1-qRT-R
ZW_P296	TCCTTTGCTACGTTGAGCCACTG	ELF6-qRT-R	**ZW_P661**	TAGTCAAGCATTGGGCTGTACTT	AGO2-qRT-F
ZW_P297	AACCTCGCATCTGGAGTGGAAC	HDA6-qRT-F	**ZW_P662**	CGAGAAGCTTCATCTATCACGGA	AGO2-qRT-R
ZW_P298	ATCTTCACCGGTAGAGTCCCTGTC	HDA6-qRT-R	**ZW_P672**	CGGCAAGCACTGAAGCTAGT	DRM1-qRT-F
ZW_P299	ACCTGGGGCGCTCAAACCTT	Sungene-qRT-F	**ZW_P673**	GAGCTCTCGGGGATTTGCTT	XLOC_024976-qRT-F
ZW_P300	CCCACTTGCCGCCATCATCCA	Sungene-qRT-R	**ZW_P674**	TGTTCTCTCTTTGCCTCCGTC	XLOC_024976-qRT-R
ZW_P301	CGACGATAAAGAATGCCGACGATG	ATLANTYS2-2-qRT-F	**ZW_P675**	ACTTCCCAAACCGTCACACT	XLOC_024441-qRT-F
ZW_P302	CGAGGAACTTGAGGGTGCGTTTAC	ATLANTYS2-2-qRT-R	**ZW_P676**	GGCTTCTCTCAGACGAAACGA	XLOC_024441-qRT-R
ZW_P303	TCTCACACAACACGGCACAGTC	ATLINE1-4-qRT-F	**ZW_P815**	TGTTGCTTCATGCTTTGTGTGT	XLOC_013763-qRT-F
ZW_P304	AGAAATCCCTCCGGTGGTGAAC	ATLINE1-4-qRT-R	**ZW_P816**	ATAACGTCGACAGCAAGCGG	XLOC_013763-qRT-R

ZW_P305	GCAAAGCACAACATGGCTGACG	AtCOPIA4-qRT-F	**ZW_P817**	TGTAAATGCATGTGCAACGGA	XLOC_025146-qRT-F
ZW_P306	ACCACCATGGAGAGTGAGCTTTG	AtCOPIA4-qRT-R	**ZW_P818**	GGCTGGCACAGCAGAAAATG	XLOC_025146-qRT-R

Table S4: Primers for T-DNA insertion lines

Primer names	Sequences (5' → 3')	Comments	Primer names	Sequences (5' → 3')	Comments
ZW_P354	ATTTTGCCGATTTCGGAAC	LBb1.3	**ZW_P481**	TGTGGAAGAGGAATTGATTGG	ago 2-1_LP
ZW_P355	ATCCATGTGGTTCAGTCTTGC	atm-LP	**ZW_P482**	AGCACCAATGAACATGACCTC	ago 2-1_RP
ZW_P356	TTGGTATCCTGCAGAGGAAAG	atm-RP	**ZW_P483**	AACTGTGTCTCGCTAACACCG	ago 7-1_a_LP
ZW_P359	CAAAGAGTCGCTTTGTTCCTG	SALK_014731-LP	**ZW_P484**	GGCAAGACCTTGTTACCTTCC	ago 7-1_a_RP
ZW_P360	TATCACTTGCCTTTTCAACGG	SALK_014731-RP	**ZW_P493**	TGTTGGTCTCCTCTGAAGCTC	jmjd5-1_LP
ZW_P477	AGATCGCTTCCAGAGTTAGCC	drm2-2_LP	**ZW_P494**	GTTCATTTATCTGCCCATTCG	jmjd5-1_RP
ZW_P478	TTGTCGCAAAAAGCAAAAGAG	drm2-2_RP	**ZW_P529**	AACCATTTTCCTCTCAGGTCG	ago4-5_LP
ZW_P479	GAGCCGTCTCATCAAACTGAC	drm1-2_LP	**ZW_P530**	GACCCTGAAGATTAGCCACATC	ago4-5_RP
ZW_P480	TTGCAGGAGCAAATATGGAAC	drm1-2_RP			

Table S5: Primers for Bisulfite sequencing analysis

Primer names	Sequences (5' →3')	Comments	Primer names	Sequences (5' →3')	Comments
ZW_P573	AATATGTTTYGGAAAATTGGAAATG	AT1TE69975-BSP2-F	ZW_P611	GTGGTATGATTAAAAAYYAYAGGTATAT	WRR4-prom.-BSP2-F
ZW_P574	TTATCATCCRRACTTCTRAATTTTT	AT1TE69975-BSP2-R	ZW_P612	CATAACAARTTRTATACTRTCCATTCTT	WRR4-prom.-BSP2-F
ZW_P597	AAAAYYAYAGGTATATATGAATATG	WRR4-prom.-BSP-F	**ZW_P613**	AGGAYAATTATTAGYAAAGTGAAGATTT	AT1TE69990-BSP1-F
ZW_P598	ACACRCCTATTRTTRTTAACATCTT	WRR4-prom.-BSP-F	**ZW_P614**	RCATATTTAACAATATCTTTATTTAACA	AT1TE69990-BSP1-R
ZW_P609	TAATTGTTTATTTTGGTYAAATATGTTT	AT1TE69975-BSP3-F	**ZW_P615**	GTATTAAAAYTGYATGTTAATGATATTT	AT1TE69990-BSP2-F
ZW_P610	ATATTTTCTAACRTTTTCTTAACATATA	AT1TE69975-BSP3-R	**ZW_P616**	CTCAATTATCCARATATCTCAATTATAA	AT1TE69990-BSP2-R

Table S6: Go enrichment comparison of upregulated genes in WT and atm mutant

GO information			Comparison		WT		*atm*	
GO Term	**Onto**	**Description**	**WT**	***atm***	**FDR**	**Num**	**FDR**	**Num**
GO:0050896	P	response to stimulus			6.1e-07	117	5,00E-05	67
GO:0010212	P	response to ionizing radiation			1.2e-06	8	7.4e-05	5
GO:0033554	P	cellular response to stress			7.1e-06	25	---	---
GO:0006950	P	response to stress			7.1e-06	74	7,00E-06	48
GO:0006259	P	DNA metabolic process			7.1e-06	25	---	---
GO:0006281	P	DNA repair			0.00017	16	---	---
GO:0006974	P	response to DNA damage stimulus			0.00022	16	---	---
GO:0042221	P	response to chemical stimulus			0.00028	63	---	---
GO:0010332	P	response to gamma radiation			0.00031	5	---	---
GO:0009404	P	toxin metabolic process			0.00052	8	0.00063	6
GO:0009407	P	toxin catabolic process			0.00052	8	0.00063	6
GO:0019748	P	secondary metabolic process			0.00084	23	2.5e-07	22
GO:0010200	P	response to chitin			0.001	12	---	---
GO:0006260	P	DNA replication			0.0028	10	---	---
GO:0019438	P	aromatic compound biosynthetic process			0.0039	14	0.012	9

GO:0009743	P	response to carbohydrate stimulus			0.0042	14	---	---
GO:0051716	P	cellular response to stimulus			0.0045	30	---	---
GO:0010035	P	response to inorganic substance			0.0048	15	---	---
GO:0010876	P	lipid localization			0.0048	5	3.5e-27	20
GO:0006302	P	double-strand break repair			0.0075	5	---	---
GO:0042435	P	indole derivative biosynthetic process			0.0097	6	---	---
GO:0042434	P	indole derivative metabolic process			0.016	6	---	---
GO:0042430	P	indole and derivative metabolic process			0.016	6	---	---
GO:0042398	P	cellular amino acid derivative biosynthetic process			0.027	12	0.011	9
GO:0007049	P	cell cycle			0.033	13	---	---
GO:0009607	P	response to biotic stimulus			0.041	22	0.00044	19
GO:0051707	P	response to other organism			0.041	21	0.013	15
GO:0003824	F	catalytic activity			2.8e-10	236	0.00024	124
GO:0017111	F	nucleoside-triphosphatase activity			5.8e-07	39	---	---
GO:0016818	F	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides			1,00E-06	39	---	---
GO:0016817	F	hydrolase activity, acting on acid anhydrides			1,00E-06	39	---	---
GO:0016462	F	pyrophosphatase activity			1,00E-06	39	---	---
GO:0016887	F	ATPase activity			1.4e-06	28	---	---
GO:0019825	F	oxygen binding			1.5e-06	20	0.024	9
GO:0030554	F	adenyl nucleotide binding			4.1e-06	53	---	---
GO:0001883	F	purine nucleoside binding			4.1e-06	53	---	---
GO:0001882	F	nucleoside binding			4.1e-06	53	---	---
GO:0017076	F	purine nucleotide binding			1.7e-05	57	---	---
GO:0016787	F	hydrolase activity			2.5e-05	94	---	---
GO:0005524	F	ATP binding			3.6e-05	48	---	---
GO:0032559	F	adenyl ribonucleotide binding			3.9e-05	48	---	---
GO:0032555	F	purine ribonucleotide binding			0.00013	52	---	---
GO:0032553	F	ribonucleotide binding			0.00013	52	---	---
GO:0004364	F	glutathione transferase activity			0.00025	8	0.0011	6
GO:0042626	F	ATPase activity, coupled to transmembrane movement of substances			0.00039	13	---	---
GO:0043492	F	ATPase activity, coupled to movement of substances			0.00039	13	---	---
GO:0016820	F	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances			0.00041	13	---	---
GO:0015399	F	primary active transmembrane transporter activity			0.00043	15	---	---
GO:0015405	F	P-P-bond-hydrolysis-driven transmembrane transporter activity			0.00043	15	---	---
GO:0000166	F	nucleotide binding			0.001	62	---	---
GO:0005488	F	binding			0.0065	221	---	---
GO:0042623	F	ATPase activity, coupled			0.0073	16	---	---
GO:0043169	F	cation binding			0.01	54	---	---
GO:0043167	F	ion binding			0.01	54	---	---

GO:0016740	F	transferase activity			0.01	78	---	---
GO:0046872	F	metal ion binding			0.01	52	---	---
GO:0022804	F	active transmembrane transporter activity			0.012	23	---	---
GO:0045735	F	nutrient reservoir activity			0.013	6	2.6e-07	10
GO:0005509	F	calcium ion binding			0.016	11	---	---
GO:0016491	F	oxidoreductase activity			0.016	40	0.00028	32
GO:0009055	F	electron carrier activity			0.02	13	0.037	9
GO:0022857	F	transmembrane transporter activity			0.05	31	---	---
GO:0006869	P	lipid transport			---	---	1.4e-11	18
GO:0033036	P	macromolecule localization			---	---	3.7e-07	21
GO:0009698	P	phenylpropanoid metabolic process			---	---	0.00044	10
GO:0009699	P	phenylpropanoid biosynthetic process			---	---	0.00044	9
GO:0006952	P	defense response			---	---	0.01	18
GO:0051704	P	multi-organism process			---	---	0.01	18
GO:0006979	P	response to oxidative stress			---	---	0.01	11
GO:0006575	P	cellular amino acid derivative metabolic process			---	---	0.02	10
GO:0051179	P	localization			---	---	0.033	31
GO:0008289	F	lipid binding			---	---	2.7e-07	18
GO:0020037	F	heme binding			---	---	0.015	6
GO:0016746	F	transferase activity, transferring acyl groups			---	---	0.029	10
GO:0016684	F	oxidoreductase activity, acting on peroxide as acceptor			---	---	0.034	6
GO:0004601	F	peroxidase activity			---	---	0.034	6
GO:0016747	F	transferase activity, transferring acyl groups other than amino-acyl groups			---	---	0.037	9
GO:0046906	F	tetrapyrrole binding			---	---	0.037	6
GO:0012505	C	endomembrane system			---	---	5.8e-10	72
GO:0031225	C	anchored to membrane			---	---	0.01	10
GO:0044464	C	cell part			---	---	0.042	162
GO:0005623	C	cell			---	---	0.042	162

P: biological process, F: molecular function, C: cellular component.

Table S7: Go enrichment comparison of downregulated genes in WT and *atm* mutant

GO information			CM		WT		*atm*	
GO Term	**Onto**	**Description**	**WT**	***atm***	**FDR**	**Num**	**FDR**	**Num**
GO:0007049	P	cell cycle			2.2e-14	34	2.7e-06	14
GO:0051726	P	regulation of cell cycle			1.4e-13	24	4.9e-06	10
GO:0042221	P	response to chemical stimulus			2,00E-13	94	3.1e-13	55
GO:0050896	P	response to stimulus			3.6e-12	141	1.9e-12	78
GO:0010033	P	response to organic substance			7,00E-09	62	7.8e-13	43
GO:0009719	P	response to endogenous stimulus			1.3e-07	51	0.00062	22
GO:0009725	P	response to hormone stimulus			1.7e-07	48	0.00022	22
GO:0009733	P	response to auxin stimulus			9.3e-07	26	---	---
GO:0007017	P	microtubule-based process			1.2e-06	15	2,00E-05	9
GO:0022402	P	cell cycle process			4.6e-06	16	---	---

GO:0007018	P	microtubule-based movement			4.7e-06	10	9.5e-05	6
GO:0006260	P	DNA replication			8,00E-06	14	---	---
GO:0009664	P	plant-type cell wall organization			4.7e-05	11	---	---
GO:0022403	P	cell cycle phase			0.00019	11	---	---
GO:0009641	P	shade avoidance			0.00034	5	---	---
GO:0006950	P	response to stress			0.00034	72	4.1e-11	54
GO:0006979	P	response to oxidative stress			0.00034	20	---	---
GO:0010035	P	response to inorganic substance			0.00037	18	---	---
GO:0065007	P	biological regulation			0.00037	113	5.7e-05	60
GO:0000279	P	M phase			0.0009	9	---	---
GO:0010564	P	regulation of cell cycle process			0.00099	6	---	---
GO:0006826	P	iron ion transport			0.0014	5	---	---
GO:0050794	P	regulation of cellular process			0.0018	92	1.2e-07	59
GO:0022607	P	cellular component assembly			0.002	16	---	---
GO:0050789	P	regulation of biological process			0.0025	98	2.6e-06	59
GO:0009628	P	response to abiotic stimulus			0.0029	48	0.0013	26
GO:0010038	P	response to metal ion			0.007	14	---	---
GO:0009314	P	response to radiation			0.0075	25	---	---
GO:0006259	P	DNA metabolic process			0.0082	19	---	---
GO:0000910	P	cytokinesis			0.0083	6	0.0006	5
GO:0009749	P	response to glucose stimulus			0.0086	5	---	---
GO:0009416	P	response to light stimulus			0.01	24	---	---
GO:0006325	P	chromatin organization			0.017	11	---	---
GO:0000278	P	mitotic cell cycle			0.017	6	---	---
GO:0034284	P	response to monosaccharide stimulus			0.022	5	---	---
GO:0009746	P	response to hexose stimulus			0.022	5	---	---
GO:0034728	P	nucleosome organization			0.023	6	---	---
GO:0006334	P	nucleosome assembly			0.023	6	---	---
GO:0006996	P	organelle organization			0.023	24	---	---
GO:0051276	P	chromosome organization			0.023	12	---	---
GO:0010556	P	regulation of macromolecule biosynthetic process			0.027	52	0.00018	33
GO:0006333	P	chromatin assembly or disassembly			0.027	7	---	---
GO:0019219	P	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process			0.027	52	0.00018	33
GO:0006261	P	DNA-dependent DNA replication			0.027	6	---	---
GO:0031497	P	chromatin assembly			0.027	6	---	---
GO:0065004	P	protein-DNA complex assembly			0.027	6	---	---
GO:0080090	P	regulation of primary metabolic process			0.028	54	0.00042	33
GO:0000041	P	transition metal ion transport			0.028	6	---	---
GO:0009723	P	response to ethylene stimulus			0.031	11	8.6e-08	14
GO:0009889	P	regulation of biosynthetic process			0.033	52	0.00022	33
GO:0031326	P	regulation of cellular biosynthetic process			0.033	52	0.00022	33
GO:0006323	P	DNA packaging			0.033	6	---	---

GO:0015674	P	di-, tri-valent inorganic cation transport			0.034	5	---	---
GO:0051171	P	regulation of nitrogen compound metabolic process			0.034	52	0.00022	33
GO:0009743	P	response to carbohydrate stimulus			0.038	12	1.2e-18	26
GO:0046686	P	response to cadmium ion			0.04	10	0.028	6
GO:0009639	P	response to red or far red light			0.04	11	---	---
GO:0060255	P	regulation of macromolecule metabolic process			0.044	55	0.00098	33
GO:0009744	P	response to sucrose stimulus			0.044	5	---	---
GO:0009753	P	response to jasmonic acid stimulus			0.045	11	0.016	7
GO:0034285	P	response to disaccharide stimulus			0.05	5	---	---
GO:0003777	F	microtubule motor activity			3,00E-11	18	1.5e-09	12
GO:0003774	F	motor activity			7.1e-10	18	1.1e-08	12
GO:0016538	F	cyclin-dependent protein kinase regulator activity			1.1e-06	10	1.8e-07	8
GO:0005199	F	structural constituent of cell wall			1.2e-06	10	---	---
GO:0019887	F	protein kinase regulator activity			4.5e-06	10	7,00E-07	8
GO:0003677	F	DNA binding			5,00E-06	92	0.00016	45
GO:0019207	F	kinase regulator activity			6.3e-06	10	9.9e-07	8
GO:0016684	F	oxidoreductase activity, acting on peroxide as acceptor			0.00029	12	---	---
GO:0016209	F	antioxidant activity			0.00029	13	---	---
GO:0004601	F	peroxidase activity			0.00029	12	---	---
GO:0030528	F	transcription regulator activity			0.00041	73	3.2e-05	42
GO:0003700	F	transcription factor activity			0.00048	67	6.3e-06	41
GO:0020037	F	heme binding			0.0036	9	---	---
GO:0009055	F	electron carrier activity			0.013	15	0.0024	10
GO:0030234	F	enzyme regulator activity			0.014	17	0.029	9
GO:0017111	F	nucleoside-triphosphatase activity			0.019	28	0.00016	20
GO:0004693	F	cyclin-dependent protein kinase activity			0.019	5	---	---
GO:0005506	F	iron ion binding			0.022	10	---	---
GO:0046906	F	tetrapyrrole binding			0.025	9	0.048	5
GO:0016818	F	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides			0.027	28	0.00024	20
GO:0016817	F	hydrolase activity, acting on acid anhydrides			0.027	28	0.00024	20
GO:0016564	F	transcription repressor activity			0.027	6	0.0029	5
GO:0015250	F	water channel activity			0.027	5	---	---
GO:0016462	F	pyrophosphatase activity			0.027	28	0.00024	20
GO:0005372	F	water transmembrane transporter activity			0.027	5	---	---
GO:0005875	C	microtubule associated complex			1.8e-07	11	0.00023	6
GO:0015630	C	microtubule cytoskeleton			1.8e-07	16	0.00036	8
GO:0044430	C	cytoskeletal part			6.8e-06	16	0.002	8
GO:0005856	C	cytoskeleton			1,00E-05	17	0.0035	8
GO:0044427	C	chromosomal part			0.00043	12	---	---
GO:0005694	C	chromosome			0.0049	12	---	---
GO:0005634	C	nucleus			0.0063	72	---	---
GO:0032993	C	protein-DNA complex			0.0063	6	---	---

GO:0005576	C	extracellular region			0.0071	15	---	---
GO:0009524	C	phragmoplast			0.011	6	0.002	5
GO:0000786	C	nucleosome			0.016	5	---	---
GO:0043232	C	intracellular non-membrane-bounded organelle			0.023	33	---	---
GO:0043228	C	non-membrane-bounded organelle			0.023	33	---	---
GO:0000785	C	chromatin			0.036	6	---	---
GO:0010200	P	response to chitin			---	---	5.8e-23	26
GO:0006952	P	defense response			---	---	3.9e-10	29
GO:0002376	P	immune system process			---	---	8.6e-08	18
GO:0006955	P	immune response			---	---	8.6e-08	18
GO:0045087	P	innate immune response			---	---	1.3e-06	16
GO:0042742	P	defense response to bacterium			---	---	7.1e-05	10
GO:0009414	P	response to water deprivation			---	---	9.3e-05	11
GO:0009873	P	ethylene mediated signaling pathway			---	---	9.5e-05	7
GO:0009415	P	response to water			---	---	0.00012	11
GO:0045449	P	regulation of transcription			---	---	0.00013	33
GO:0006915	P	apoptosis			---	---	0.00017	9
GO:0006350	P	transcription			---	---	0.00017	34
GO:0002252	P	immune effector process			---	---	0.00019	5
GO:0000160	P	two-component signal transduction system (phosphorelay)			---	---	0.00022	7
GO:0009737	P	response to abscisic acid stimulus			---	---	0.00022	13
GO:0009617	P	response to bacterium			---	---	0.00058	10
GO:0010468	P	regulation of gene expression			---	---	0.00062	33
GO:0009651	P	response to salt stress			---	---	0.00063	12
GO:0031323	P	regulation of cellular metabolic process			---	---	0.00081	33
GO:0051707	P	response to other organism			---	---	0.0013	15
GO:0016265	P	death			---	---	0.0015	10
GO:0006970	P	response to osmotic stress			---	---	0.0015	12
GO:0008219	P	cell death			---	---	0.0015	10
GO:0012501	P	programmed cell death			---	---	0.0021	9
GO:0051301	P	cell division			---	---	0.0022	6
GO:0009607	P	response to biotic stimulus			---	---	0.0022	15
GO:0019222	P	regulation of metabolic process			---	---	0.0029	33
GO:0006355	P	regulation of transcription, DNA-dependent			---	---	0.0037	19
GO:0009620	P	response to fungus			---	---	0.0037	7
GO:0043687	P	post-translational protein modification			---	---	0.0039	22
GO:0051252	P	regulation of RNA metabolic process			---	---	0.0039	19
GO:0016567	P	protein ubiquitination			---	---	0.0054	6
GO:0006351	P	transcription, DNA-dependent			---	---	0.0061	19
GO:0032774	P	RNA biosynthetic process			---	---	0.0061	19
GO:0009987	P	cellular process			---	---	0.0061	116
GO:0032446	P	protein modification by small protein conjugation			---	---	0.009	6
GO:0009611	P	response to wounding			---	---	0.011	7

GO:0009751	P	response to salicylic acid stimulus			---	---	0.012	7
GO:0007010	P	cytoskeleton organization			---	---	0.012	6
GO:0051704	P	multi-organism process			---	---	0.012	15
GO:0006464	P	protein modification process			---	---	0.012	23
GO:0009409	P	response to cold			---	---	0.012	9
GO:0007165	P	signal transduction			---	---	0.015	20
GO:0009266	P	response to temperature stimulus			---	---	0.016	11
GO:0070647	P	protein modification by small protein conjugation or removal			---	---	0.016	6
GO:0050832	P	defense response to fungus			---	---	0.017	5
GO:0009755	P	hormone-mediated signaling pathway			---	---	0.036	8
GO:0032870	P	cellular response to hormone stimulus			---	---	0.036	8
GO:0043412	P	macromolecule modification			---	---	0.038	23
GO:0007167	P	enzyme linked receptor protein signaling pathway			---	---	0.046	5
GO:0007169	P	transmembrane receptor protein tyrosine kinase signaling pathway			---	---	0.046	5
GO:0004888	F	transmembrane receptor activity			---	---	5.5e-05	10
GO:0004872	F	receptor activity			---	---	0.00024	10
GO:0016563	F	transcription activator activity			---	---	0.0019	8
GO:0060089	F	molecular transducer activity			---	---	0.0024	12
GO:0004871	F	signal transducer activity			---	---	0.0024	12
GO:0005488	F	binding			---	---	0.0043	114
GO:0005516	F	calmodulin binding			---	---	0.0078	8
GO:0005524	F	ATP binding			---	---	0.016	22
GO:0032559	F	adenyl ribonucleotide binding			---	---	0.016	22
GO:0030554	F	adenyl nucleotide binding			---	---	0.028	22
GO:0001882	F	nucleoside binding			---	---	0.028	22
GO:0001883	F	purine nucleoside binding			---	---	0.028	22
GO:0032555	F	purine ribonucleotide binding			---	---	0.044	23
GO:0032553	F	ribonucleotide binding			---	---	0.044	23
GO:0031224	C	intrinsic to membrane			---	---	0.002	18

P: biological process, F: molecular function, C: cellular component.

Table S8: Regulated TEs/repeat elements in WT

TEs with coding potential

	ID	Log2FC	Families/Types	TAIR ids	Length	old Name
TEs: 28	AT3TE83365	7.17	LTR/Copia	AT3G55300 copia	1248	ATCOPIA37
	AT2TE32120	5.57	LINE/L1	AT2G17910	4440	TA11
	AT5TE82660	3.43	Copia	AT5G56830	729	Copia
	AT5TE45925	3.23	LTR/Gypsy	AT5G34480	4597	ATGP8
	AT1TE52220	2.82	DNA/MuDR	AT1G42420	6302	VANDAL4
	AT2TE28025	2.78	LTR/Gypsy	AT2G15810	544	ATGP10
	AT3TE50020	2.32	DNA/En-Spm	AT3G30393	1004	ATENSPM5
	AT4TE56270	2.14	LTR/Copia	AT4G23160,overlap	2163	ATCOPIA11
	AT1TE80020	1.74	LINE/L1	AT1G65485	498	ATLINE2

	AT1TE72060	1.67	LTR/Copia	AT1G59265 copia	4814	ATRE1
	AT5TE25615	1.24	retrotransposon	AT5G20880	1123	-
	AT1TE22850	-1.04	LTR/Gypsy	AT1G20390	4440	ATLANTYS2
	AT4TE67490	-1.12	LTR/Copia	AT4G28900	5006	ATCOPIA46
	AT5TE36920	-1.47	LTR/Copia	AT5G28145	5103	ATCOPIA12
	AT2TE24530	-1.49	LTR/Copia	AT2G14245	1742	ATCOPIA77
	AT2TE24525	-1.50	DNA/En-Spm	AT2G14230	4129	ATENSPM1A
	AT2TE54780	-1.55	LTR/Copia	AT2G29165 copia	2385	ATCOPIA63
	AT5TE71590	-1.59	LTR/Gypsy	AT5G49080	651	ATHILA6A
	AT4TE21055	-1.63	LTR/Gypsy	AT4G08109	13366	ATHILA0_I
	AT5TE49725	-1.65	LTR/Copia	AT5G35777	4774	ATCOPIA12
	AT1TE39880	-1.70	LTR/Copia	contains AT1G33813 and AT1G33817 two copia elements	3129	ATCOPIA35
	AT5TE71595	-1.85	LTR/Gypsy	part of AT5G49080, EXT11 transposon	657	ATHILA6A
	AT5TE71590	-1.87	LTR/Gypsy	AT5G49080	651	ATHILA6A
	AT1TE57585	-2.01	LINE/L1	AT1G47405	711	ATLINEIII
	AT1TE57025	-2.44	LTR/Copia	AT1G46120	2406	ATCOPIA63
	AT1TE34650	-2.60	LINE/L1	AT1G30390	1786	ATLINEIII
	AT2TE79235	-2.92	DNA/Harbinger	AT2G42050 unknow type TE	1049	SIMPLEGUY1
	AT5TE37660	-3.26	LTR/Gypsy	AT5G28335 gypsy	791	ATGP3

Non coding RNAs derived from TEs, Repeats and other regions

	ID	Log2FC	Families/Types	TAIR ids	Length	old Name
Relics of TEs: 48	AT1TE38195	9.28	RC/Helitron	-	937	ATREP15
	AT1TE38190	7.71	RC/Helitron	-	193	ATREP14
	AT4TE29715	5.92	RC/Helitron	-	626	ATREP2A
	AT2TE42460	5.48	LINE?	-	626	TSCL
	AT4TE22775	5.34	LTR/Copia	-	450	ATCOPIA37
	AT5TE59075	4.60	RC/Helitron	-	3168	ATREP3
	AT3TE47515	4.07	LTR/Copia	-	598	ATCOPIA37
	AT2TE42055	4.02	RC/Helitron	-	793	HELITRONY1B
	AT2TE19695	3.58	RC/Helitron	-	973	ATREP3
	AT1TE69990	3.55	RC/Helitron	-	840	ATREP3
	AT2TE13060	3.49	LTR/Copia	-	620	ATCOPIA32
	AT1TE04725	3.47	DNA/MULE-MuDR	-	169	ATDNAI27T9B
	AT3TE43415	3.14	DNA/hAT-Ac	-	486	ATHATN1
	AT3TE43200	3.10	LTR/Copia	-	592	ATCOPIA37
	AT1TE51920	3.00	DNA/MULE-MuDR	-	431	ATDNAI27T9C
	AT3TE59970	2.93	DNA/MULE-MuDR	-	560	ATDNAI27T9B
	AT3TE41165	2.87	RC/Helitron	-	1500	ATREP3

	AT5TE93870	2.87	Satellite	-	596	HELITRONY1D
	AT1TE71780	2.83	LTR/Copia	-	153	ATCOPIA67
	AT1TE04720	2.73	DNA/MULE-MuDR	-	410	ATDNAI27T9C
	AT1TE04710	2.66	DNA/MULE-MuDR	-	663	ATDNAI27T9B
	AT3TE48430	2.51	RC/Helitron	-	1221	ATREP11
	AT1TE04700	2.50	DNA/MULE-MuDR	-	131	ATDNAI27T9A
	AT4TE84220	2.43	DNA/hAT-Ac	-	557	ATHATN6
	AT5TE65325	2.40	DNA/MULE-MuDR	-	874	VANDAL22
	AT1TE64430	1.97	LTR/Copia	-	299	ATCOPIA57
	AT5TE93875	1.94	Satellite	-	664	ATREP11
	AT4TE84225	1.89	DNA/MULE-MuDR	-	1767	ARNOLDY1
	AT1TE29390	1.82	DNA/CMC-EnSpm	-	1126	HELITRONY3
	AT5TE55640	1.77	LINE/L1	-	691	ATLINE1_3A
	AT5TE29590	1.71	DNA/PIF-Harbinger	-	923	SIMPLEGUY1
	AT4TE09225	1.50	LTR/Copia	-	1107	ATCOPIA37
	AT2TE26610	-1.51	LTR/Gypsy	-	2068	ATGP2N
	AT4TE33910	-1.60	DNA	-	607	TNAT2A
	AT3TE00350	-1.64	LTR/Copia	-	212	ATCOPIA24
	AT5TE21595	-1.66	DNA/MULE-MuDR	-	1591	ARNOLDY1
	AT5TE37060	-1.96	RC/Helitron	-	3359	HELITRONY3A
	AT3TE08290	-2.06	SINE	-	278	RathE2_cons
	AT5TE64390	-2.41	LTR/Gypsy	-	242	VANDAL16
	AT4TE12660	-2.43	LINE/L1	-	298	TA11
	AT1TE75910	-2.48	LINE/L1	-	226	ATLINE1A
	AT3TE08270	-2.59	SINE	-	278	RathE2_cons
	AT1TE16325	-3.19	DNA	-	698	TNAT1A
	AT1TE03900	-3.76	RC/Helitron	-	822	ATREP10D
	AT3TE89880	-3.81	LTR/Copia	-	381	ATCOPIA41
	AT5TE39275	-4.07	LINE/L1	-	807	ATLINE1_6
	AT2TE47790	-5.43	DNA/hAT-Ac	-	463	TAG2
	AT5TE71740	-5.65	RC/Helitron	-	1308	ATREP3
Repeats:	AT5TE08075	7.30	Simple_repeat	-	220	VANDAL1
22	AT1TE64490	4.45	Low_complexity	-	499	ATREP15
	AT1TE69975	3.96	Simple_repeat	-	876	ATREP4
	AT4TE03175	3.77	Simple_repeat	-	428	BRODYAGA1A
	AT5TE55815	3.37	Simple_repeat	-	372	ATDNA2T9C
	AT1TE30070	3.11	Simple_repeat	-	966	ATREP3
	AT4TE45160	2.73	Simple_repeat	-	323	ATDNA2T9C
	AT5TE29585	2.02	Simple_repeat	-	1164	ATREP5
	AT2TE45275	1.95	Simple_repeat	-	667	ATREP9
	AT1TE36140	1.88	Simple_repeat	-	139	ATCOPIA27

	AT1TE74530	1.72	Simple_repeat	-	1350	ATREP10D
	AT3TE02100	1.66	Simple_repeat	-	403	VANDAL17
	AT5TE47605	-1.75	Simple_repeat	-	329	ATHILA6A
	AT2TE45520	-1.96	Simple_repeat	-	352	ATHILA6A
	AT4TE23045	-2.11	Simple_repeat	-	956	ATREP10D
	AT1TE04990	-2.21	Simple_repeat	-	164	ATHAT10
	AT1TE15090	-2.50	Simple_repeat	-	1481	ATREP10D
	AT1TE71325	-3.13	Simple_repeat	-	439	ATREP10D
	AT5TE91030	-3.16	Low_complexity	-	418	ATHAT10
	AT2TE01730	-5.18	Simple_repeat	-	342	ATREP11
	AT5TE27520	-5.31	Simple_repeat	-	758	ATREP15
	AT5TE59610	-5.61	Low_complexity	-	990	ATREP9
Just ncRNAs : 18	AT1TE38180	8.46	-	-	155	ATREP11
	AT2TE00090	5.69	-	-	279	TAT1_ATH
	AT5TE25580	4.76	-	-	634	VANDAL8
	AT1TE69970	4.11	-	-	149	ATREP3
	AT5TE37695	3.42	-	-	264	VANDAL6
	AT3TE48425	3.02	-	-	179	ATREP6
	AT3TE44530	2.70	-	-	303	ATENSPM5
	AT3TE44395	2.15	-	-	211	VANDAL21
	AT2TE57405	2.07	-	-	407	VANDAL8
	AT3TE43770	1.52	-	-	429	HELITRON4
	AT2TE18115	-1.46	-	-	174	ATGP2
	AT5TE21300	-1.70	-	-	381	HELITRONY3
	AT4TE22225	-2.18	-	-	163	ATHILA6A
	AT1TE13245	-2.51	-	-	187	VANDAL1
	AT1TE42665	-3.12	-	-	299	ATREP15
	AT2TE28180	-3.27	-	-	159	VANDAL1
	AT5TE70115	-4.12	-	-	562	HELITRON4
	AT4TE10810	-5.70	-	-	137	ATHATN1

Table S9: Regulated TEs/repeat elements in *atm*

TEs with coding potential

	ID	Log2FC	Families/Types	TAIR ids	Length	old Name
TE: 1	AT2TE07145	1.37	AT2G04460	-	4373	ATCOPIA95

Non coding RNAs derived from TEs, Repeats and other regions

Relic of TE: 1	AT1TE07180	-5.30	DNA/MULE-MuDR	-	373	VANDAL5A
Repeat: 1	AT5TE08075	4.62	Simple_repeat	-	220	VANDAL1
Just ncRNAs : 1	AT1TE83905	7.17	-	-	231	HELITRONY1E

Table S10: LncRNAs regulated by X-ray irradiation and co-expressed genes in WT

ID	Log2 FC	TAIR annotated	PLncDB	CPC prediction	Co-expression genes	Log2 FC	Gene name	Position relate to gene	ncRNA Type (PlncDB annotation)	Length of lncRNAs (bp)	WEB repeatmasker
XLOC_024976	7.22		RLFS_026432		-	-	-	-	lincRNA	668	
XLOC_024441	3.64		At4NC060340		AT4G30530	0.73	*GGP1*	5' of gene	lincRNA	844	
XLOC_009836	3.50	AT2G30362			AT2G30360	3.91	*CIPK11*	Overlap	NAT	2747	
XLOC_006240	2.55		RLFS_006415		-	-	-	-	lincRNA	1692	
XLOC_008153	2.20		RLFS_007748		-	-	-	-	RCTU	1329	no repetitive seq
XLOC_008145	1.92		At1NC064450		-	-	-	-	RCTU	587	Simple_repeat, A rich
XLOC_027207	1.60		At5NC066580		AT5G42600	1.79	*MRN1*	5' of gene	RCTU	1878	Simple_repeat
XLOC_019295	1.27		RLFS_021173		-	-	-	-	GATU/lincRNA	1040	
XLOC_010202	1.22	AT2G36792			AT2G36800	1.39	*DOGT1*	between AT2G36780 and AT2G36800	NAT	1775	
					AT2G36780	5.10					
XLOC_016117	1.01	AT3G52748			-	-	-	-	lincRNA	2233	
XLOC_009679	-1.75		At2NC048830		-	-	-	-	RCTU	1299	no repetitive seq
XLOC_019840	-2.33	AT3G61198			-	-	-	-	lincRNA	717	
XLOC_013763	-2.58		At3NC007260		AT3G06435	-1.35	-	3' of gene	RCTU	800	Simple_repeat
AT1TE38195	9.28			noncoding	AT1G32570	1.30		3' of gene		937	RC/Helitron
AT5TE08075	7.30		At5NC007680		-	-	-	-		220	Simple_repeat
AT4TE29715	5.92			noncoding	-	-	-	4th intron of AT1G11340		626	RC/Helitron
AT2TE00090	5.69			noncoding						279	-
AT2TE42460	5.48		At2NC041940,At2		-	-	-	-		626	LINE?

		NC041950							
AT4TE22775	5.34	At4NC024710		-	-	-	-	450	LTR/Copia
AT5TE25580	4.76		noncoding	AT5G20850	5.35	*ATRAD51*	overlap with 3' of gene	634	-
AT5TE59075	4.60		noncoding	AT5G40840	4.42	*SYN2*	3' of gene	3168	RC/Helitron
AT1TE64490	4.45	part of Group1183-NATs		AT1G52315	1.03		3' of gene	499	Low_complexity
AT3TE47515	4.07		noncoding	-	-	-	-	598	LTR/Copia
AT2TE42055	4.02		noncoding	-	-	-	-	793	RC/Helitron
AT1TE69975	3.96		noncoding	AT1G56510	1.29	*WRR4*	intron 1	876	Simple_repeat
AT4TE03175	3.77		noncoding	AT4G01450	1.34	*UMAMIT30*	5' of gene	428	Simple_repeat
AT2TE19695	3.58		noncoding	AT2G11810	2.29	*MGDC*	overlap with 3' of gene	973	RC/Helitron
AT1TE69990	3.55		noncoding	AT1G56510	1.29	*WRR4*	intron 4	840	RC/Helitron
AT2TE13060	3.49		noncoding	-	-	-	-	620	LTR/Copia
AT5TE37695	3.42		noncoding	-	-	-	-	264	-
AT5TE55815	3.37	At5NC060460		-	-	-	-	372	Simple_repeat
AT5TE45925	3.23	At5NC050160		-	-	-	-	4597	LTR/Gypsy
AT3TE43415	3.14		noncoding	-	-	-	-	486	DNA/hAT-Ac
AT1TE30070	3.11	RLFS_008432,At1NC033500		-	-	-	-	966	Simple_repeat
AT3TE43200	3.10	Seed_Group3214		-	-	-	-	592	LTR/Copia
AT1TE51920	3.00		noncoding	-	-	-	-	431	DNA/MULE-MuDR
AT3TE59970	2.93	At3NC058680, cover partial		-	-	-	-	560	DNA/MULE-MuDR

		AT3TE59960 and AT3TE59970, so this two TE elements can be actual one TE							
AT3TE41165	2.87		noncoding	-	-	-	-	1500	RC/Helitron
AT5TE93870	2.87	RLFS_033687		-	-	-	-	596	Satellite
AT1TE52220	2.82	At1NC058150		-	-	-	-	6302	DNA/MuDR
AT4TE45160	2.73	At4NC041990		AT4G17905	1.79	*ATL4H*	3' of gene	323	Simple_repeat
AT1TE04720	2.73	At1NC005250, RLFS_008357		-	-	-	-	410	DNA/MULE-MuDR
AT3TE44530	2.70		noncoding	AT3G28580	2.11		in 3' of gene	303	-
AT1TE04710	2.66		noncoding	-	-	-	-	663	DNA/MULE-MuDR
AT3TE48430	2.51		noncoding	-	-	-	-	1221	RC/Helitron
AT4TE84220	2.43	At4NC069570		AT4G37022 AT4G37030	2.97 2.97		in between of AT4G37022 and AT4G37030	557	DNA/hAT-Ac
AT5TE65325	2.40		noncoding	-	-	-	-	874	DNA/MULE-MuDR
AT3TE44395	2.15		noncoding	AT3G28510	3.08	-	3' of gene	211	-
AT2TE57405	2.07		noncoding	AT2G30750	2.17	*CYP71A12*	overlap with 3' of gene	407	-
AT5TE29585	2.02	At5NC030160,At5NC030170		AT5G24150 AT5G24155	1.23	*SQP1*	in between of AT5G24150 and AT5G24155	1164	Simple_repeat

AT1TE64430	1.97		noncoding	AT1G52270	2.29	-	overlap with 3' of gene	299	LTR/Copia
AT2TE45275	1.95	AT2G24755		-	-	-	-	667	Simple_repeat
AT5TE93875	1.94	RLFS_033687		-	-	-	-	664	Satellite
AT4TE84225	1.89		noncoding	AT4G37022 AT4G37030	2.97		in between of AT4G37022 and AT4G37030	1767	DNA/MULE-MuDR
AT1TE29390	1.82	Group764		AT1G26380	1.76	-	overlap with 3' of gene	1126	DNA/CMC-EnSpm
AT5TE55640	1.77		noncoding	-	-	-	-	691	LINE/L1
AT1TE74530	1.72		noncoding	AT1G61260	1.06	-	overlap with 3' of gene	1350	Simple_repeat
AT5TE29590	1.71	RLFS_033934		AT5G24150 AT5G24155	1.23	*SQP1*	in between of AT5G24150 and AT5G24155	923	DNA/PIF-Harbinger
AT3TE02100	1.66	Group3357		AT3G02400	2.39		in 3' of gene	403	Simple_repeat
AT3TE43770	1.52		noncoding	AT3G28270	1.64		in 5' of gene	429	-
AT4TE09225	1.50	Group4932, AT4G03935,NATs		-	-	-	-	1107	LTR/Copia
AT2TE24530	-1.49	At2NC026520		-	-	-	-	1742	LTR/Copia
AT2TE26610	-1.51	At2NC028610		AT2G15042	-1.19	-	overlap with 5' of gene	2068	LTR/Gypsy
AT4TE33910	-1.60		noncoding	-	-	-	-	607	DNA
AT4TE21055	-1.63	RLFS_026290, RLFS_026291		-	-	-	-	13366	LTR/Gypsy

		and others							
AT3TE00350	-1.64		noncoding (weak)	AT3G01260	-1.51		in 5' of gene	212	LTR/Copia
AT5TE21595	-1.66		noncoding (weak)	AT5G18030	-1.28		3' of gene	1591	DNA/MULE-MuDR
AT5TE21300	-1.70		noncoding					381	-
AT5TE47605	-1.75		noncoding (weak)					329	Simple_repeat
AT5TE37060	-1.96	RLFS_027873						3359	RC/Helitron
AT2TE45520	-1.96		noncoding (weak)	AT2G24980	-1.66	*EXT6*	in 3' of gene	352	Simple_repeat
AT3TE08290	-2.06		noncoding					278	SINE
AT4TE23045	-2.11		noncoding					956	Simple_repeat
AT5TE64390	-2.41		noncoding	AT5G44260	-2.66	-	3' of gene	242	LTR/Gypsy
AT4TE12660	-2.43	At4NC011260 and others		-	-	-	-	298	LINE/L1
AT1TE75910	-2.48	RLFS_003260		-	-	-	-	226	LINE/L1
AT1TE15090	-2.50	At1NC016670, At1NC016680		-	-	-	-	1481	Simple_repeat
AT3TE08270	-2.59	At3NC007250		AT3G06435	-1.35	-	3' of gene	278	SINE
AT1TE42665	-3.12	RLFS_008472		-	-	-	-	299	-
AT1TE71325	-3.13	RLFS_007115		-	-	-	-	439	Simple_repeat
AT5TE91030	-3.16		noncoding	AT5G63090	-0.97	LOB	3' of gene	418	Low_complexity
AT1TE16325	-3.19	At1NC018170		-	-	-	-	698	DNA
AT1TE03900	-3.76	AT1G04425 NAT		-	-	-	-	822	RC/Helitron

AT3TE89880	-3.81	ATCOPIA41 NAT contains TE		-	-	-	-	381	LTR/Copia
AT5TE39275	-4.07	Group6774 NAT		AT5G28770	-1.02	*ZIP63*	5' of gene	807	LINE/L1
AT5TE70115	-4.12		noncoding	AT5G48070	-3.47	*XTH20*	overlap with 3' of gene	562	-
AT2TE01730	-5.18		noncoding	-	-	-	-	342	Simple_repeat
AT5TE27520	-5.31	At5NC028060		-	-	-	-	758	Simple_repeat
AT2TE47790	-5.43		noncoding (weak)	-	-	-	-	463	DNA/hAT-Ac
AT5TE59610	-5.61		noncoding	-	-	-	-	990	Low_complexity
AT5TE71740	-5.65		noncoding	-	-	-	-	1308	RC/Helitron

lincRNA: long intergenic non-coding RNA. NAT: natural antisense transcript. RCTU: repeat-containing transcription unit. GATU: gene-associated transcription unit.

Table S11: LncRNAs regulated by X-ray irradiation and co-expressed genes in *atm* mutant

	Log2 FC	**TAIR annotated**	**PLncDB**	**CPC prediction**	**Co-expression genes**	**Log2FC**	**Gene name**	**Position relate to gene**	**ncRNA Type**	**Length of lncRNAs (bp)**	**WB repeatmasker**
XLOC_024976	5.11		RLFS_026432		-	-	-	-	lincRNA	668	
XLOC_024441	2.58		At4NC060340		-	-	-	-	lincRNA	844	
XLOC_027207	2.24		At5NC066580		AT5G42600	4.12	*MRN1*	5' of gene	RCTU	1878	
XLOC_008153	1.86		RLFS_007748		-	-	-	-	RCTU	1329	simple_repeat
XLOC_019295	1.81		RLFS_021173		-	-	-	-	GATU	1040	
XLOC_018207	1.58	AT3G27884			-	-	-	-	GATU	2133	
XLOC_028948	1.34	AT5G07322			AT5G07330	2.31	-	5' of gene	LincRNA	446	
XLOC_009836	1.11	AT2G30362			-	-	-	-	NAT	2747	

XLOC_028191	-1.23	AT5G59732			AT5G59730	-1.36	*EXO70H7*	NAT of AT5G59730	NAT	2295	
XLOC_019748	-1.39	AT3G59765			-	-	-	-	RCTU	1884	
XLOC_013612	-1.56		RLFS_014451		-	-	-	-	GATU	1317	
XLOC_000180	-1.67	AT1G04425			-	-	-	-	RCTU	2873	
XLOC_032116	-1.87		RLFS_033703		-	-	-	-	RCTU	1124	
XLOC_025146	-3.28		RLFS_026616		-	-	-	-	RCTU	952	RC/Helitron, ATREP10A
XLOC_002820	-3.69	AT1G56242			AT1G56240	-4.00	*AtPP2-B13*	NAT of AT1G56240	NAT	1585	
					AT1G56250	-3.30	*AtPP2-B14*	3' of gene			
AT1TE83905	7.17		Seed_Group1223		-	-	-	-		231	
AT5TE08075	4.62		At5NC007680		-	-	-	-		220	
AT1TE07180	-5.30			noncoding	AT1G07160	-4.51	-	in 5' of gene		373	

lincRNA: long intergenic non-coding RNA. NAT: natural antisense transcript. RCTU: repeat-containing transcription unit. GATU: gene-associated transcription unit.