The genetic Characterization of RNA-directed Transcriptional Gene Silencing in *Arabidopsis thaliana*

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I

- Biowissenschaften -

der Martin-Luther-Universität Halle-Wittenberg

von Andreas Finke

geboren am 11.09.1981 in Wolfen

„In der Wissenschaft gleichen wir alle nur den Kindern, die am Rande des Wissens hier und da einen Kiesel aufheben, während sich der weite Ozean des Unbekannten vor unseren Augen erstreckt."

Sir Isaac Newton (1643 – 1727)

Acknowledgment

I would like to thank Dr. Michael Florian Mette for giving me the opportunity to be part of the epigenetics group, continuous support, guidance, helpful suggestions invaluable effort and patience in correcting this thesis.

I thank the members of the research group of Epigenetics Markus Kuhlmann, Che How Teo and Branimira Borisova Todorova for helpful discussions, suggestions, scientific advices and for keeping a nice working atmosphere in the lab.

Furthermore I want to express my gratitude to Christa, Beate and Inge for their excellent technical support.

I am very thankful to Dr. Renate Schmidt, for helpful discussions and suggestions.

I want to express my gratitude to Ines Walde and Susanne König for their excellent work in the IPK sequencing facility and for many cheerful discussions.

Moreover I want to thank Martin Mascher for analyzing the next generation sequencing data.

I also thank Prof. Gunter Reuter, Dr. habil. Ortrun Mittelsten Scheid and Dr. Michael Florian Mette, for their acceptance to be part of the evaluation committee for my PhD work.

I thanks to all colleagues outside of the Epigenetics group that helped to keep my time in the IPK in good remembrance.

Mein allergrößtes Dankeschön gilt meiner Familie, meine Eltern Albrecht und Christine, meine Schwester Marianne und meine Freundin Alexandra die mich während der Zeit meiner Doktorarbeit unterstützt und immer wieder motiviert haben.

Ein besonderes Dankeschön geht darüber hinaus an Alice Mosch für die hilfreichen Gespräche zu (fast) jeder Tages und Nachtzeit.

CONTENT

Main Figures

Supplementary Figures

LIST OF TABLES

Main Tables

Supplementary Tables

Abbreviations

°C	degree centigrade	min	minute(s)
3' 35S	3' UTR of 35S	ml	milliliter
3' DRM2	3'UTR of DRM2	mM	millimolar
3' IDN2	3' UTR of IDN2	mRNA	messenger RNA
3' nos	3'UTR of *NOPALIN SYNTHASE*	MS	Murashige & Skoog
3' ocs	3' UTR of OCTOPINE SYNTHASE	NGS	Next generation sequencing
A. thaliana	*Arabidopsis thaliana*	*nrd*	no RNA-directed transcriptional gene silencing
A. tumefaciens	*Agrobacterium tumefaciens*		
approx.	approximately	nt	nucleotide
AtMU1	*A. thaliana* Mutator-like 1	OD_{600}	Opical density at 600 nm wave length
AtSN1	*A. thaliana* SINE 1	ORF	open reading frame
bp	Base pair	p4-RNA	Pol IV dependent RNA
C_1; C_2; C_3	1st / 2nd / 3rd generation of non-mutagenized contol plants	p5-RNA	Pol V dependent RNA
		PCR	Polyymerase chain reaction
DNA	Deoxyribonucleic acid	*Pro35S*	35S promoter of Cauliflower mosaic virus
dNTP	deoxynucleoside triphosphate	*ProDRM2*	Promoter of *DRM2*
DTT	Dithiotreithol	*ProIDN2*	Promoter of *IDN2*
E. coli	*Escherichia coli*	*ProMAS*	Promoter of *MANNOPINE SYNTHASE* of *A. tumefaciens*
EDTA	Ethylenediaminetetraacetic acid		
ELISA	Enzyme-linked immunosorbent assay	*ProNOS*	Promoter of *NOPALINE SYNTHASE* of *A. tumefaciens*
et al.	*et alii*		
FPN1	Ferroportin 1	RdDM	RNA-directed DNA methylation
g	gram	RdTGS	RNA-directed transcriptional gene silencing
GM	germination medium	RNA	ribonucleic acid
GUS	ß-glucuronidase	RNAi	RNA interference
h	hour(s)	rpm	rounds per minute
H3	Histone 3	RT	room temperature
H4	Histone 4	RT-PCR	reverse transcription PCR
H3K9me2	Histone 3 dimethylated at lysine 9	SDS	Sodium dodecylsulfat
HELICc	Helicase C-terminal domain	sec	second(s)
HMG-Box	High Mobility Group Box	SINE	small interspersed nuclear element
HPT	*HYGROMYCIN PHOSPHOTRANSFERASE*	siRNA	small interferring RNA
Hyg	hygromycin	SNP	single nucleotide polymorphism
Hyg^R	Hygromycin resistant	SOB	Super Optimal Broth
IGN	Intergenic noncoding locus	SOC	Super Optimal Broth + Glucose
K4	Lysine 4	SSC	Sodium Chloride Sodium Citrate Buffer
K9	Lysine 9	SWI2/SNF2	SWIch2/Sucrose non-fermenting
K20	Lysine 20	T_1; T_2	1st / 2nd generation of plants after transformation
K27	Lysine 27		
K36	Lysine 36	TBE	Tris-Borat-EDTA
Kan	kanamycin	TE	transposable element
Kan^R	Kanamycin resistant	TIR	terminal inverted repeat
Kan^S	Kanamycin sensitiv	UTR	untranslated region
kb	kilo base pairs	x *g*	times gravity of earth
LB	Lysogeny broth	YEB	yeast exact broth
lncRNA	long non-coding RNA	F_1; F_2; F_3	1st / 2nd / 3rd filial generation
M	Molarity		
M_1;M_2; M_3	1st / 2nd / 3rd generation of plants after mutagenesis		

1 INTRODUCTION

1.1 EPIGENETICS

Organisms are subjected to various developmental processes throughout their life cycle. The proper confinement of these processes requires a reliable temporal and spatial regulation of gene expression. Nevertheless, the relevant executable information, the genes encoded by the storage medium DNA, usually stays essentially the same in every cell of an individual. This also seems reasonable, as developmental regulation based on DNA sequence change would face the problem that coordinated emergence of identical alterations (mutations) in the primary information in multiple cells would be statistically unlikely. Hence, the differential execution of genetic information in differentiated cells needs to be regulated in other ways. One option is *via* the regulation of the accessibility of particular parts of information, that is, particular regions of the DNA.

Extensive compaction of the DNA double strand into higher order structures is necessary to facilitate the incorporation of genomic DNA into the eukaryote nucleus. The first level of compaction happens at nucleosomes, the basic repeating unit of eukaryotic chromatin (Kornberg and Klug, 1981) consisting of an octamer core particle, linker histone H1 and DNA. The core particle is formed from two copies each of the four major histone proteins H2A, H2B, H3 and H4, around which 147 bp of superhelical DNA double strand are wrapped. Individual core particles are separated by 20 to 80 bp of DNA, to which H1 is bound during interphase. The wrapping of the DNA around the core particles leads to the formation of a beads-on-a-string structure, which can be observed electron-microscopically under non-physiological conditions. In living cells, this string structure is supposed to form a 30 nm fiber structure, which can undergoes further steps of compaction to finally form the highly condensed chromosomes present in metaphase (Van Holde, 1988; Bednar *et al.,* 1998).

The whole entity of DNA and its associated proteins in the eukaryote nucleus is termed chromatin. In addition to the compaction of the nuclear DNA, a further important function of chromatin is the selective regulation gene expression in the context of cell differentiation. Chromatin-mediated transcriptional regulation can be mitotically and in some cases meiotically stable, that is, heritable, without the involvement of any changes

of DNA sequence. The deciphering of the mechanisms regulating the access to genetic information is the main scope of epigenetic research.

Piloting cytogenetic experiments using light microscopy led to the definition of two major chromatin states, the relatively more condensed heterochromatin and the more open euchromatin. Heterochromatin can be further divided into constitutive heterochromatin, which is late replicating and remains condensed during the cell cycle in virtually all cells of an organism, and facultative heterochromatin, which is formed from euchromatin in a cell type- or condition-specific manner. In the model plant *Arabidopsis thaliana* (*A. thaliana*), the cytogenetically defined constitutive heterochromatin is located in and around the centromeres. It is enriched in repetitive sequences and silenced transposable elements (TEs), but contains only few expressed genes, and therefore is considered to be a low expressing chromatin compartment. In contrast, euchromatin located on the chromosome arms is rich in highly expressed genes and depleted in TEs, and thus is referred to as an actively expressing chromatin compartment.

1.2   MECHANISM OF EPIGENETIC REGULATION

Histone Modifications

The histones H2A, H2B, H3 and H4 forming the nucleosome core are highly basic proteins. While their C-terminal domains arrange into a defined globular structure that binds to the minor grove of DNA, their N-terminal domains are rather unstructured and protrude from the core particle. It is well established that these N-terminal "tails" are subjected to diverse post-translational modifications. Of the many covalent histone modifications known, methylation, acetylation, phosphorylation, sumoylation and ribosylation were so far identified in plants (Berr *et al.,* 2011; Dong *et al.,* 2012; Houben *et al.,* 2007; Luo *et al.,* 2008; Dhawan *et al.,* 2009; Miller *et al.,* 2010; Bannister and Kouzarides 2011). For most of these, the way how the modification alters chromatin structure is not well understood. It is assumed that they affect nucleosome-nucleosome interactions, nucleosome-DNA interactions and / or the interaction of nucleosomes with non-histone proteins and by this regulate transcription and other processes at chromatin level.

Variation in chromatin structure can also be brought about by the incorporation of histone variants. In many eukaryotes, histone proteins are encoded by gene families with functional differentiation of family members. For example, in *A. thaliana* histone variant H3.1 is enriched in silent and variant H3.3 is associated with actively transcribed

chromatin (Stroud *et al.,* 2012). Similarly, the H2A variant H2A.Z marks regions around gene promoters (Zilberman *et al.,* 2008; March-Diáz and Reyes, 2009), and cenH3, which is a more divergent form of histone H3, is located exclusively in centromeric regions, where it plays essential functions including proper segregation of chromosomes (Lermantova *et al.,* 2006).

Methylation of lysine (K) and arginine (R) residues can happen to different levels, as mono-, di- and trimethylation at K and mono- and dimethylation at R, which further increases the diversity of possible histone modification patterns (Naumann *et al.,* 2005; Cloos *et al.,* 2008). The high number of possible combinations of these modifications and their assumed interdependence led to the formulation of the histone code hypothesis, which postulates that the combination of different histone modification at a chromatin region leads to a certain consistent regulatory outcome (Turner 2000; Strahl and Allis. 2001; Jenuwein and Allis, 2001).

Among the known histone modifications, methylation and acetylation of lysine residues located in the N-terminal domains of histone H3 and H4 have been most extensively studied and are thought to be associated with transcriptional activation or repression.

Recently, integrative analyses of histone modification profiles performed in *A. thaliana* revealed the organization of the "epigenome" into four major chromatin states. Actively expressed genes show enrichment for di- and trimethylation of K4 and K36 of H3, while repressed genes within euchromatin are marked by trimethylation of K27 of H3. Furthermore, silenced TEs are marked by monomethylation of K20 of H4 and dimethylation of K9 of H3, while intergenic regions and low expressed genes do not display any prevalent marks (Roudier *et al.,* 2011).

DNA methylation

In eukaryotes DNA methylation refers commonly to the enzymatic transfer of a methyl group ($-CH_3$) to the cyclic carbon 5 in the pyrimidine ring of cytosine. In contrast to mammals, where methylation in somatic cells is predominantly limited to the cytosines in CG context, DNA methylation in plants is found in CG, CHG and CHH (with H standing for C, A, T) contexts (Ramsahoye *et al.,* 2000; Lister *et al.,* 2009; Cokus *et al.,* 2008). In *A. thaliana*, approximately 7% of the cytosines in cellular DNA are found to be methylated; with 24%, 6.7% and 1.7% of cytosines in CG, CHG and CHH context methylated, respectively (Rohzon *et al.,* 2009; Cokus *et al.,* 2008). Generally, DNA methylation in plants is a mark of heterochromatin and transcriptional inactivation that is

typically associated with a silenced chromatin state and is largely confined to silent repetitive sequences and TEs.

Repression of transposable elements

As mentioned above, in *A. thaliana* the majority of DNA methylation is located in pericentromeric regions of chromosomes, which mainly consist of TEs and other repetitive sequences. Excessive expression of TEs would present a threat to host genome integrity, as frequent random integration would lead to a high incidence of gene disruption. Therefore, repression of TE transcription is necessary. Analysis of TE inactivation in maize first pointed to the importance of DNA methylation in TE silencing (Chandler and Walbot, 1986; Banks *et al.,* 1988). Later on, the loss of DNA methylation in *A. thaliana* was found to cause decondensation of pericentromeric regions, which is accompanied by transcriptional reactivation of previously silenced TEs (Miura *et al.,* 2001; Lippman *et al.,* 2003; Zhang *et al.,* 2006; Mathieu *et al.,* 2007; Blevins *et al.,* 2009). Further, in a more physiological context, the removal of DNA methylation by DNA glycosylase DEMETER (DME) in the vegetative nucleus of pollen in *A. thaliana* causes the transcriptional activation of diverse types of TEs (Slotkin *et al.,* 2009; Hsieh *et al.,* 2009; Gehring *et al.,* 2009).

Parental Imprinting

DME is also involved in parental imprinting in *A. thaliana*, a mechanism leading to the parent-of-origin specific expression of genes during embryogenesis and seed development. Among the few imprinted genes known in *A. thaliana*, the regulation of *MEDEA* (*MEA*), *FLOWERING WAGENINGEN* (*FWA*), *PHERES1* and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*) are best described (Grossniklaus *et al.,* 1998; Luo *et al.,* 1999; Kinoshita *et al.,* 1999; Kinoshita *et al.,* 2004; Ville-Calzada *et al.,* 1999; Köhler *et al.,* 2005). *MEA*, *FWA* and *FIS2* are only active in the central cell of the female gametophyte (Choi *et al.,* 2002; Kinoshita *et al.,* 2004). After double fertilization, only the maternal alleles are expressed in the endosperm. While inactivity of paternal *FWA* and *FIS2* in endosperm and of both alleles of the two genes in all other tissues seems to depend solely on DNA hypermethylation of direct sequence repeats in upstream of their respective promoters (Kinoshita *et al.,* 2004; Jullien *et al.,* 2006a), silencing of the paternal and maternal alleles of *MEA* and *PHE1*, respectively, seems to involve additional mechanisms despite a clear correlation between DNA methylation in the promoter and transcriptional silencing in *MEA* (Xiao *et al.,* 2003). At these gene loci, histone-modifying Polycomb complexes are required for stable silencing. Silencing at

MEA (Jullien *et al*, 2006b) as well as at the repressed maternal *PHE1* (Köhler et al., 2005) rather depends on repressive histone marks. Recently, a region in the MEA locus was identified that is essential for transcriptional silencing, but is not differentially methylated in endosperm and vegetative tissue, which further supports the assumption of a DNA methylation independent imprinting mechanism (Wöhrmann *et al.,* 2012).

Paramutation

Paramutation is an epigenetic mechanism that refers to an *in trans* inactivation between homologous alleles leading to reproducible and heritable changes in gene expression at one of the alleles (Stam and Mittelsten Scheid, 2005). Paramutation has been mainly studied in maize, but paramutation and paramutation-like phenomena were also observed in other plant species as well as in mice and human (Krebbers *et al.,* 1987; Hagemann *et al.,* 1993; Bennett *et al.,* 1997;). Although the precise underlying mechanism is still enigmatic, a number of structural, genetic and epigenetic requirements for establishment of the silenced state could be identified. A feature found in many paramutation systems are direct or inverted repeat sequences which are, at least in cases like the locus *booster1*, indispensable for the paramutation process (Kermicle *et al.,* 1995; English and Jones, 1998; Walker and Panavas, 2001; Stam *et al.,* 2002; Sidorenko and Chandler, 2008). Furthermore, based on the observation that unsilenced paramutable alleles display low levels of DNA methylation and become methylated and silenced when combined with a paramutagenic allele, an involvement of DNA methylation in paramutation was assumed (Walker and Panavas, 2001; Haring *et al.,* 2010;). This was further supported by forward genetic screens in maize that identified homologs of *A. thaliana* NRPD1, NRPD2a, RDR2 and a SWI2/SNF2-like factor to be essential for paramutation, suggesting the involvement of a RNA-directed DNA methylation-like mechanism (see below) (Dorweiler *et al.,* 2000; Alleman *et al.,* 2006; Hale *et al.,* 2007; Erhard *et al.,* 2009; Sidorenko *et al.,* 2009; Stonaker *et al.,* 2009; Law *et al.,* 2011).

1.3   ESTABLISHMENT AND MAINTENANCE OF DNA METHYLATION

DNA methyltransferases

Cytosine methylation is conferred to DNA post-replicatory by a set of DNA methyltransferases (DMTase). All eukaryotic DMTases identified so far share a number of common amino acid motives in their respective methyltransferase domains (MTase), which are involved in recognition of the target site, binding of the co-factor S-adenosyl

methionine and in catalyzing the transfer of the methyl group (Posfai *et al.,* 1989; Kumar *et al.,* 1994; Goll and Bestor, 2005). As these sequence motives are also conserved in the DNA methyltransferases of bacteria, eukaryotic and bacterial DMTases share a common evolutionary root.
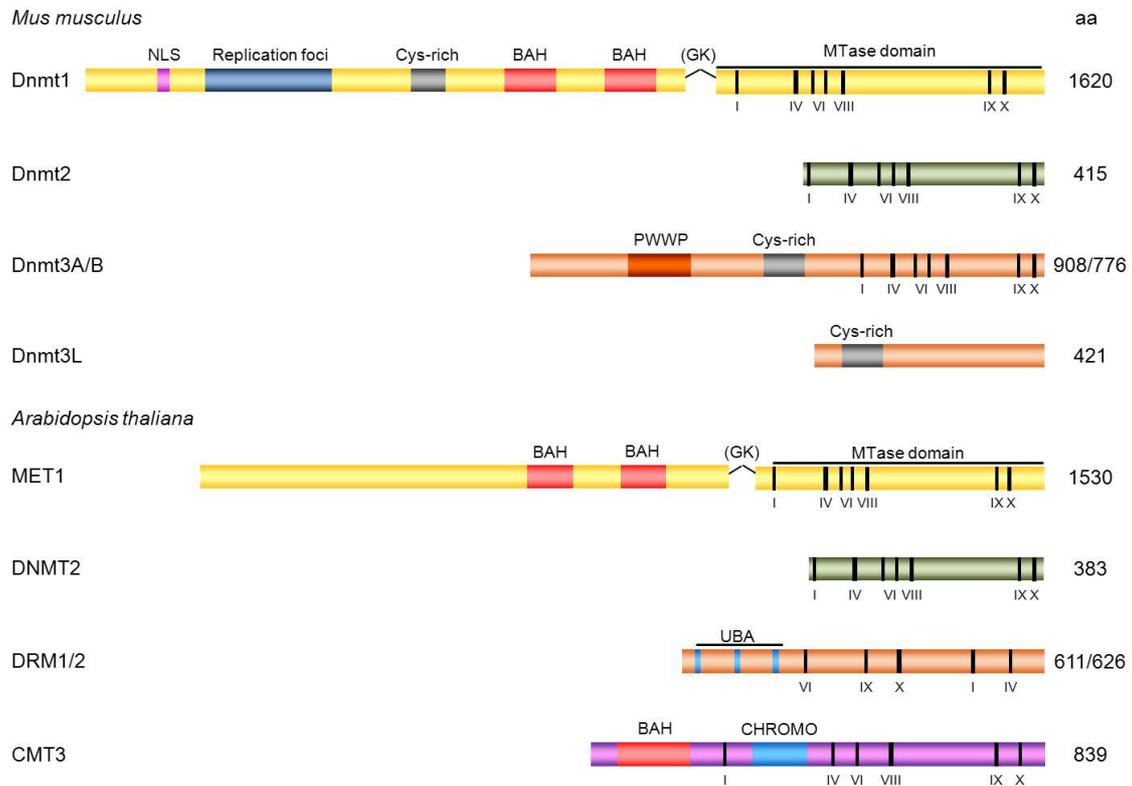
Figure 1: Classes and Domain organisation of DNA methyltransferases of *M. musculus* and *A. thaliana*.
Positions and order of conserved sequence motifs in the MTase domain are indicated. Figure 1 was adapted from Goll and Bestor 2005.

Phylogenetic studies comparing the MTase domains of eukaryotic DNMTases led to the definition of four families (Figure 1). While homologs of three families, Dnmt1, Dnmt2 and Dnmt3, have been identified in plant and non-plant species, the chromodomain-containing CHROMOMETHYLASE family is plant specific (Henikoff and Comai, 1998).

In *A. thaliana*, 10 genes for proteins displaying sequence homology to MTase domains were identified (Finnegan and Kovac, 2000). Four putative MTases, termed *METHYLTRANSFERASE 1* (*MET1*) and *METHYLTRANSFERASE1-LIKE* (*MEL*), belong to the *Dnmt1* family. *DOMAINS REARRANGED METYHLTRANSFERASE* (*DRM*) *1* to *3* are homologs of *Dnmt3* and *CHROMOMETHYLASE* (*CMT*) *1* to *3* are the members of the plant specific DMTase family. In addition, one *Dnmt2*-homolog is encoded in the *A. thaliana* genome, which is only very lowly expressed and to which no function in genomic DNA methylation is designated to.

In *A. thaliana*, like in mammals, a stark correlation of methylation in one strand with methylation on the opposite strand exists for strand-symmetric CG sites (Bird, 2002; Cokus *et al.,* 2008). The re-establishment of full CG context methylation on hemimethylated DNA double strand after DNA replication depends on the activity of DMTase MET1, SWItch2/Sucrose non fermenting 2 (SWI2/SNF2) chromatin remodeling factor DECREASE IN DNA METHYLATION (DDM1) and VARIANT IN METHYLATION (VIM) proteins 1 to 3 (Vongs *et al.,* 1993; Woo *et al.,* 2007; Woo *et al.,* 2008). Furthermore, the interaction of MET1 with HISTONE DEACETYLASE 6 (HDA6) is necessary to maintain CG context methylation at certain loci such as the centromeric 180 bp repeats (Aufsatz *et al.,* 2002b; To *et al.,* 2011; Liu *et al.,* 2012). The molecular mechanisms underlying the CG context methylation maintenance machinery is poorly understood in plants; however, the existence of proteins analogous to MET1, VIMs and DDM1 in mammals might suggests that similar mechanisms are acting in mammals and in plants (Law and Jacobsen, 2010).

Like in the case of the CG context, the positioning of the cytosines in CHG context is symmetric on the two strands of double stranded DNA. Consistently, a stark correlation of methylation in one strand with the methylation in the opposite strand was reported for the CHG context, which would suggest the action of a maintenance mechanism similar to the one for CG methylation maintenance (Cokus *et al.,* 2008). However, the maintenance of methylation in CHG context differs significantly from CG methylation maintenance. The majority of CHG methylation in *A. thaliana* depends on the DMTase CHROMOMETHYLASE3 (CMT3) and on H3K9-specific SUPPRESSOR OF VARIAGATION3-9 HOMOLOG (SUVH) histone methyltransferases SUVH4, SUVH5, SUVH6 (Jackson *et al.,* 2002; Malagnac *et al.,* 2002; Ebbs *et al.,* 2006; Pontvianne *et al.,* 2012). Based on early *in vitro* experiments, which showed affinity of the chromodomain of CMT3 to peptides dimethylated at K9of histone H3 and the decrease of CHG methylation in H3K9me2-deficient mutants, it was concluded that these histone marks serve to label CMT3 methylation target sites (Lindroth *et al.,* 2004). Recently, this interpretation was confirmed by a genome-wide superimposition of CMT3 binding sites and H3K9me2-marked nucleosomes (Du *et al.,* 2012). Moreover, the involvement of HDA6 in CHG methylation maintenance at several loci is also well documented (Aufsatz *et al.,* 2002; Liu *et al.,* 2012; To *et al.,* 2011).

In contrast to CG and CHG context, the cytosines in CHH context do not have a symmetrically positioned counterpart on the complementing DNA strand. Thus, the

mechanism for the propagation of methylation in CHH context needs to be more complex and involves persistent *de novo* methylation by DRM-type DNA methyltransferases. In contrast to mammalian *de novo* methyltransferases Dnmt3a and Dnmt3b, the MTase domain of DRM-type enzymes displays a rearranged order of catalytic motives (Figure 1; Cao *et al.,* 2001). Furthermore, DRM proteins contain three UBA domains (UBIQUITIN ASSOCIATED) able to bind ubiquitin, which are supposed to direct DRMs to the sites of *de novo* methylation (Mueller and Feigon, 2002; Kozlov *et al.,* 2007; Henderson *et al.,* 2010). Although DRMs are encoded by three genes in *A. thaliana*, only *DRM1* and *DRM2* encode proteins that contain all invariant residues necessary for catalytic activity (Cao *et al.,* 2001; Cao and Jacobsen 2002; Julien *et al.,* 2012). In contrast, *DRM3* is not expected to display catalytic activity due to the absence of the invariant cytidylprolyl dipeptide of motiv IV, which is thought to be necessary for recognition and stabilization of the cytosine in the active center and an additional mutation of conserved residues in motive IX and X (Henderson *et al.,* 2010, Bestor and Verdine, 1994).

Sequence specificity is conferred to DRM by an RNA interference (RNAi)-related mechanism known as RNA-directed DNA methylation (RdDM) first observed in *Nicotiana tabacum* infected with potato spindle tuber viroid (Wassenegger *et al.,* 1994). The critical requirement for RdDM is formation of double-stranded (ds)RNA, was demonstrated by Pol II-mediated transcription of inverted repeat (IR) structures. The resulting transcript with partial self-complementarity can fold intramolecularly to form dsRNA and can efficiently trigger DNA hypermethylation of homologous sequences *in trans*. If the transcribed IRs contain promoter sequences, hypermethylation of homologous promoters is observed, which can result in stable transcriptional silencing of the affected gene(s) (Mette *et al.,* 2000; Aufsatz *et al.,* 2002a). This observation led to the design of diverse RdDM-reporter systems making use of natural occurring as well as engineered inverted repeats for forward and reverse genetic screens for "silencing suppressor" mutations (Aufsatz *et al.,* 2002b, Fischer *et al.,* 2008, Brosnan and Voinet, 2010, Finke *et al.,* 2012a, Finke *et al.,* 2012ba, Eun *et al.,* 2012)

The RdDM pathway at endogenous target sequences is assumed to be a circular, self-perpetuating mechanisms depending on the transcriptional activity of two plant specific DNA-DEPENDENT RNA POLYMERASES (Pol), Pol IV and Pol V. (Herr *et al.,* 2005; Onodera *et al.,* 2005; Kanno *et al.,* 2005; Pontier *et al.,* 2005).

RdDM is thought to be initiated upon the synthesis of single-stranded long non-coding RNAs (lncRNA) by a complex consisting of Pol IV, a 12-subunit DNA-dependent RNA

polymerase and accessory proteins SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), REDUCED IN DNA METHYLATION 4 (RDM4), an *A. thaliana* homolog of yeast Iwr1 and the SWI2/SNF2 chromatin remodeling factor CLASSY1 (CLSY1) (Smith *et al.,* 2007; He *et al.,* 2009; Kanno *et al.,* 2009; Law *et al.,* 2011).
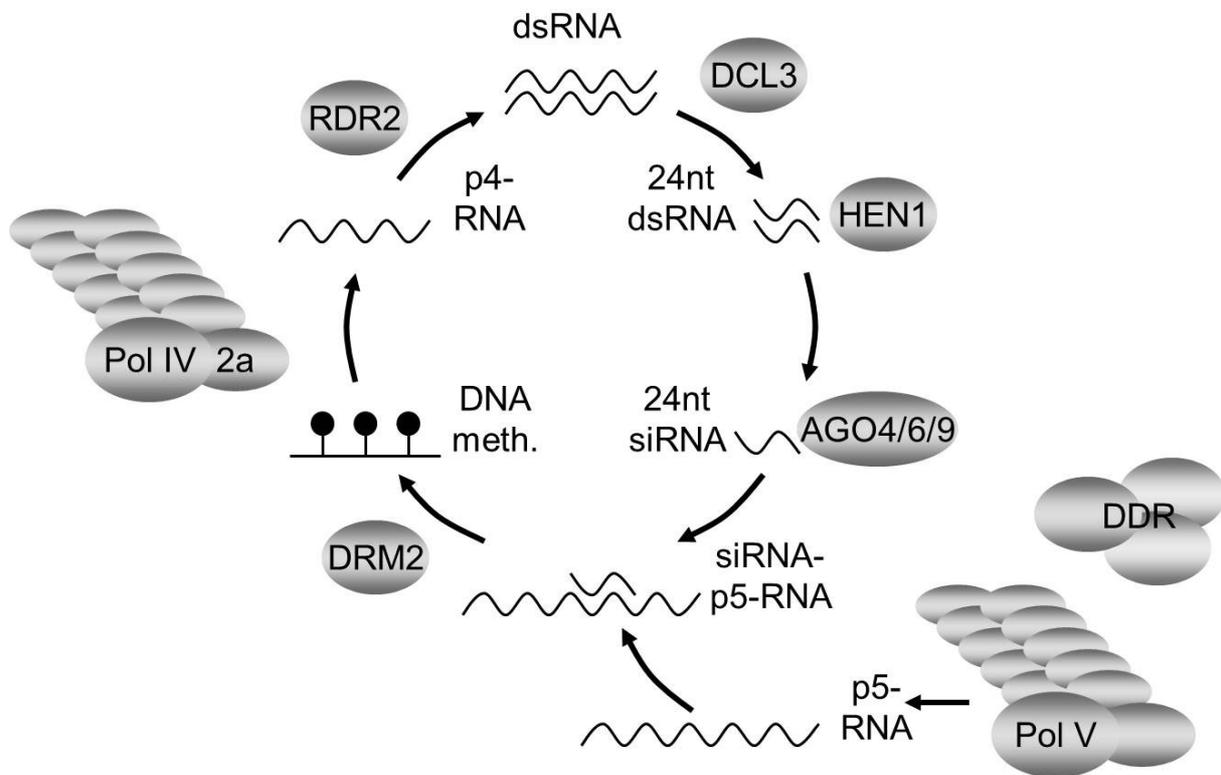
Figure 2: Core components of the RNA-directed DNA methylation pathway

These p4-RNAs are converted into long dsRNAs by RNA DEPENDENT RNA POLYMERASE 2 (RDR2) (Xie *et al.,* 2004; Chan *et al.,* 2004; Law *et al.,* 2011). The dsRNA is then processed by RNaseIII-like endonuclease DICER-LIKE 3 (DCL3) into fragments of 24 nt length with 2 nt 3'-overhangs (Xie *et al.,* 2004; Chan *et al.,* 2004). After 2'-O-methylation of the 3' terminal nucleotides by HUA ENHANCER 1 (HEN1) to prevent polyuridylation and thus, presumably, degradation (Chen *et al.,* 2002; Yu *et al.,* 2005, He *et al.,* 2009, Li *et al.,* 2005; Abe *et al.,* 2010), one of the two strands is incorporated into an ARGONAUT (AGO) protein of the AGO4-clade of *A. thaliana* (AGO4, AGO6, AGO9) to form a silencing effector complex that is recruited to chromatin. AGO4, AGO6 and AGO9 display preferential binding of 24nt siRNAs with an 5' A (Zilberman *et al.,* 2003; Zheng *et al.,* 2007; Vaucheret, 2008; Takeda *et al.,* 2008; Havecker *et al.,* 2010, Eun *et al.,* 2011).

DNA methylation *via* the RdDM pathway further requires activity of Pol V which synthesizes a further class of lncRNAs (Wierzbicki *et al.,* 2008; Wierzbicki *et al.,* 2009;

Wierzbicki *et al.,* 2012). The synthesis of these p5-RNAs depends on the SNF2/SWI2 chromatin remodeling factor DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), the SMC-like protein DEFICIENT IN MERISTEM SILENCING 3 (DMS3)/INVOLVED IN DE NOVO 1 (IDN1) and the small, single stranded DNA-binding protein REDUCED IN DNA METHYLATION 1 (RDM1) (Kanno *et al.,* 2004; Kanno *et al.,* 2005; Kanno *et al.,* 2008; He *et al.,* 2009; Ausin *et al.,* 2009; Gao *et al.,* 2010). These proteins are the known constituents of the DDR complex, which interacts with Pol V *via* binding of DRD1 and DMS3 to its largest subunit NRPE1 and presumably recruits Pol V to chromatin (Wierzbicki *et al.,* 2008; Wierzbicki *et al.,* 2009; Law *et al.,* 2011). Recently, DEFECTIVE IN MERISTEM SILENCING 11 (DMS11)/ARABIDOPSIS MICROCHIDIA 6 (AtMORC6) and AtMORC1 were identified as further factors in RdDM suppressor screens. Due to its interaction with DMS3 *in vitro*, DMS11/AtMORC6 is supposed to be part of the the DDR complex. (Lorkovic *et al.,* 2012; Moissiard *et al.,* 2012).

The sequence-specific recruitment of the siRNA-AGO4 complex to the DNA involves base pairing of the siRNA with p5-RNA. Furthermore, protein-protein interactions of AGO4 with SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L) / KOW DOMAIN CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) and the C-terminus of the large subunit of Pol V, NRPE1, are necessary (Wierzbicki *et al.,* 2008; Wierzbicki *et al.,* 2009; Bies-Etheve *et al.,* 2009, He *et al.,* 2009; Huang *et al.,* 2009). This interaction depends on the presence of the recently defined glycine-tryptophane/tryptophane-glycine (GW/WG) motif (also known as AGO-hook) (El-Shami *et al.,* 2007; Azevedo *et al.,* 2011; Karlowski *et al.,* 2010). The loss of p5-transcripts, as observed in Pol V deficient *nrpe1* mutants, does affect the localization of the siRNA-AGO complex and local DNA methylation patterns, but does not severely decrease the overall genome-scale DNA methylation level. Instead, DNA methylation is relocated, indicating a guiding rather than catalytically activating function of p5-RNAs in RdDM (Wierzbicki *et al.,* 2012). INOLVED IN DE NOVO 2 (IDN2) as well as FACTOR OF DNA METHYLATION (FDM) 1 to 5 form a family of SUPPRESSOR OF GENE SILENCING 3 (SGS3)-like RdDM factors able to bind dsRNA with 5' overhang. By this, they are implicated in stabilizing the siRNA:p5-RNA duplex (Ausin *et al.,* 2009; Zheng *et al.,* 2010; Xie *et al.,* 2012; Zhang *et al.,* 2012, Finke *et al.,* 2012b). DRM2, and in early developmental stages DRM1, are recruited to chromatin by an unresolved mechanism to confer DNA methylation, which finally causes transcriptional gene silencing in a yet unidentified manner. In analogy to the essential role of SRA domain containing VIM proteins and SUVH4/SUVH5/SUVH6 in the maintenance of CG and CHG context methylation, respectively, propagation of CHH

methylation depends on the partially redundant activity of SUVH2 and SUVH9 (Johnson *et al.,* 2008; Kuhlmann *et al.,* 2012).

1.4 DNA METHYLATION AND THEN…?

While significant progress has been made in deciphering the pathways leading to DNA methylation, the downstream mechanisms responsible for the recognition and interpretation of methylation that finally lead to transcriptional suppression are poorly understood. However, few proteins potentially involved were identified.

Mutations in MORPHEUS MOLECULE 1 (MOM1), a multimerizing CMM domain-containing protein evolved from a CHD3-like (CHROMODOMAIN-HELICASE-DNA binding) SNF2 chromatin remodeling factor, are accompanied by a release of transcriptional repression. DNA methylation is not or only very slightly reduced at loci released in *mom1* (Woodage *et al.,* 1997; Amedeo *et al.,* 2000; Steimer *et al.,* 2000; Habu *et al.,* 2006, Vaillant *et al.,* 2006 Caikovski *et al.,* 2008; Habu *et al.,* 2010; Yokthongwattana *et al.,* 2010; Nishimura *et al.,* 2012). Moreover, release of transcription of gene *SDC* in *mom1* is accompanied by reduced amounts of H3K9me2 (Numa *et al.,* 2010). Although the precise molecular function of MOM1 is still elusive, these results imply an activity downstream of the establishment of DNA methylation. Like MOM1, PICKLE (PKL) is a CHD3-like protein which is involved in transcriptional repression during embryogenesis and in meristematic tissue (Eshed *et al.,* 1999; Ogas *et al.,* 1999; Perruc *et al.,* 2007). Caikovski *et al.,* (2008) observed an additive accumulation of TE-derived *TSI* transcripts in *mom1 pkl* double mutants, pointing to a partially redundant role of both proteins.

Further components that might confer silencing after DNA methylation are methyl-cytosine binding proteins that contain a SRA or a methylcytosine-binding domain (MBD). While the involvement of SRA-domain containing proteins in DNA methylation-dependent silencing was shown in genetic screens, the role of MBD-proteins in *A. thaliana* is less well understood. MBD proteins are capable to bind DNA methylated in CG and CHH context and their interaction with DDM1 and histone deacetylases, as well as the co-localization with chromocenters and 5s rDNA clusters have been shown (Ito *et al.,* 2003; Zemach and Grafi, 2003; Schebba *et al.,* 2003).

1.5 THE ROLE OF DNA METHYLATION IN PHYSIOLOGICAL PROCESSES

Beside its involvement in epigenetic effects in the strict sense, that is, in setting up mitotically and meiotically heritable differential states of gene expression in phenomena such as TE silencing, imprinting and paramutation, dynamic DNA methylation might play a role in the reaction of plants to environmental factors. Recent studies point to vital functions of DNA methylation in the defense against pathogens. For example, *A. thaliana* plants deficient for functional *Pol V, AGO4, DRD1, RDR2, DRM2, DRD1 and MET1* were found to have altered responses to pathogenes *Pseudomonas syringae DC3000*, *Botrytis cinerea* and *Plectosphaerella cucumerina* (Agorio and Vera, 2007; Lopez *et al.,* 2011). This is in agreement with altered DNA methylation patterns observed in promoters of pathogen-responsive gene before and after pathogen exposure (Lopez and Vera, 2007; Dowen *et al.,* 2012). Furthermore, important functions of DNA methylation mechanisms in response to different abiotic stresses such as heat, high salinity and mutagenic compounds were detected in several studies. For example, activation of the TE *AtCOPIA78/ONSEN* by heat treatment persisted longer in *nrpe1* mutant plants than in wild type individuals and only in the progeny of *nrpe1* plants new *AtCOPIA78/ONSEN* insertions were observed (Ito *et al.,* 2011). Moreover, Trigger *et al.,* (2012) observed the RdDM-regulated transcriptional inactivation of genes important for stomata development as response to low relative humidity. Also, an altered response to DNA-damaging agent methyl methanesulfonate and high salt concentrations was reported for *ddm1* mutants (Yao *et al.,* 2012). The analysis of recombination during meiosis in three complementing studies revealed that in *ddm1* and *met1* plants, recombination frequencies in general stayed similar. However, the distribution of recombination points changed, leading to elevated recombination in pericentromeric regions (Melamed-Bessudo and Levy, 2012; Mirouze *et al.,* 2012; Yelina *et al.,* 2012). This indicates a function of DNA methylation in the regulation of recombination as well.

1.6 SCOPE OF THIS THESIS

The aim of this work was the isolation, identification and genetic characterization of new genes involved in the RdTGS pathway of *A. thaliana* (RdTGS factors) to further refine the knowledge about the mechanisms responsible for RdTGS. To achieve this genes, a transgenic silencing reporter system was used. In this system, the constitutive transcription of an inverted repeat (IR) of the *NOPALINE SYNTHASE* promoter (*ProNOS*) sequence (referred to as *ProNOS-IR*) in the *SILENCER* (*H*) transgene confers siRNA-dependent *in trans* hypermethylation of an unlinked *ProNOS* copy in an *TARGET* (*K*) transgene (referred to as *TARGET-ProNOS*). This hypermethylation causes transcriptional inactivation of a *NEOMYCINE PHOSPHOTRANSFERASE II* gene (*NPTII*) under the control of the *ProNOS*, rendering *H* and *K*positive plants sensitive to the aminoglycoside antibiotic kanamycin (Kan^S).

To isolate new RdTGS factors a forward genetic screen was performed. The M_2 populations of ethylmethan sulfonate (EMS) treated seeds, homozygous for both transgenes were screened for plants that show a kanamycin resistance (Kan^R) phenotype. The obtained lines should be analyzed in respect to the NPTII protein amounts, the DNA methylation of the *TARGET-ProNOS* as well as of several endogenous sequences. Finally, the mutated gene loci causative for the release of RdTGS should be identified by map based cloning.

In a previously performed similar screen, several copies of the targeted T-DNA-construct were present in the genome (Aufsatz et al., 2002). By contrast, the release of silencing of a single copy *TARGET* T-DNA insertion particularly susceptible to TGS induced by the *ProNOS-IR* (Fischer *et al.,* 2008) was assayed in the screen performed during work for this thesis.

2 METHODS AND MATERIALS

2.1 CHEMICALS

If not mentioned otherwise, chemicals and devices used in the work for this thesis were purchased from the companies Bio-Rad Laboratories GmbH (Munich, GER), Ducheefa Biochemie (Haarleem, NL), Eppendorf AG (Hamburg, GER), Carl Roth GmbH + Co KG (Karlsruhe, GER), Abimed GmbH (Langenfeld, GER), Heinemann Labortechnik GmbH (Duderstadt, GER) and Sartorius AG (Göttingen, GER). Buffers and solutions were prepared in bi-distilled water according to standard protocols (Sambrook and Russel, 2001) and autoclaved if necessary. The majority of enzymes were purchased from Fermentas/ThermoFisher (St. Leon Roth, GER), Taq Polymerase was purchased from Promega (Mannheim, GER). Radioactive labeled compounds were purchased from Hartmann Analytic (Braunschweig, GER).

2.2 PLANT MATERIAL AND GROWTH CONDIDTIONS

The *A. thaliana* accessions Columbia-0 (Col-0) and Landsberg *erecta* (L*er*) were used in all experiments. The transgenic *A. thaliana* line double homozygous for *TARGET* and *SILENCER* transgenes ($K_{chr1-10}/K_{chr1-10}$;*H/H*) has been described previously (Fischer *et al.,* 2008).

For generation of rosette leaf material for molecular analysis *A. thaliana* was cultivated on IPK greenhouse soil mixture (heat treated at 55°C over night) at 21°C under a 16 h light / 8 h dark (long day) regime for propagation and seed production; and at 21°C under a 8 h light / 16 h dark (short day) regime For antibiotics resistance tests, seeds were surface-sterilized and germinated under long day regime in environmental growth chambers (Percival Scientific Inc., CLF Laborgeräte, Emersacker, GER) on agar-plates with germination medium (GM) containing appropriate antibiotics. Seeds were stratified in darkness at 4°C for 48 – 72 h before transfer to growth chambers.

2.3 MEDIA

Table 1: Composition of media used for *in vitro* cultivation of *Arabidiopsis thaliana*

Medium	Component	per 1L	final conc.
Germination medium (GM) pH 5.7	Murashige & Skoog Salts	2.15 g	0.215 % (w/v)
	succrose	10 g	1 % (w/v)
	Fe-EDTA	5 ml (200x)	1x
	B5 vitamins	10 ml (100x)	1x
	MES	0.5 g	0.05 % (w/v)
	Bacto-Agar	8 g	0.8 % (w/v)
GM + Kan[200]	kanamycin	4 ml (50 g/l)	200 mg/l
GM + Kan[200]Hyg[20]	kanamycin	4ml (50 g/l)	200 mg/l
	hygromycin	0,4 ml (50 g/l)	20 mg/l

Table 2: Infiltration medium for *A. thaliana* transformation

Medium	Component	per 1L	final conc.
Infiltration medium	Murashige & Skoog Salts	2.15 g	0.215 % (w/v)
	sucrose	50 g	5 %(w/v)
	B5 vitamins	10 ml	1 % (w/v)
	BAP	4,25 µl (1 mg/ml)	187 nM
	Acetosyringone	1 ml (100 mg/ml)	0.1 % (w/v)
	Silwet L-77 (Lehle Seeds, USA)	250 µl	0.25 % (v/v)

Table 3: Composition of media used for bacteria cultivation

Medium	Component	per 1L	final conc.
LB medium (Bertani 1951)	Tryptone	10 g	1 % (w/v)
	Yeast extract	5 g	0.5 % (w/v)
	NaCl	10 g	170 mM
for Agar plates	Micro Agar	15 g	1.5 % (w/v)
SOB medium (Hanahan 1983)	Tryptone	20 g	2 % (w/v)
	Yeast extract	5 g	0.5 % (w/v)
	NaCl	0.5 g	170 mM
	KCl	10 ml (250 mM)	2.5 mM
	$MgCl_2$	5 ml (2 M)	10 mM
	$MgSO_4$	5 ml (2 M)	10 mM
SOC medium (Hanahan *1983*)	SOB medium + glucose	20 ml (1 M)	20 mM
YEB medium (Miller 1972)	Beef extract	5 g	0.5 % (w/v)
	Yeast extract	2 g	0.2 % (w/v)
	Peptone	5 g	0.5 % (w/v)
	Sucrose	5 g	0.5 % (w/v)
	$MgCl_2 * 6H_2O$	0.5 g	2 mM
for agar plates	Micro Agar	15 g	1,5 % (w/v)

2.4 BACTERIAL STRAINS AND VECTORS

Table 4: Bacteria Strains

Strain	Genotype	Reference
E.coli DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–*	Meselson and Yuan, 1968
E.coli Strataclone Solopack	*F⁻ Cre+ Φ80lacZΔM15 endA1 recA1 tonA* (otherwise unknown)	Agilent Technologies
Agrobacterium tumefaciens	*pGV2260* in *C58C1*	Deblaere *et al.,* 1985

Table 5: Vectors

Vector	Use	Resitance	Reference/Source
pSC-A-amp/kan	*E. coli* cloning vector	Amp^R Kan^R	Agilent Technologies
pGEM-7Zf(-)	*E. coli* cloning vector, source of MCS for *pCMBL2*	Amp^R	Promega
pCMBL2	basic binary vector	BASTA^R (plant) Kan^R (bacterial)	Finke *et al.,* 2012b and this work
pCMBL2+ ProIDN2:IDN2	complementation of *nrd1*	BASTA^R (plant) Kan^R (bacterial)	Finke et al., 2012b and this work
pCMBL2+ ProDRM2:DRM2	complementation of *nrd3-1*	BASTA^R (plant) Kan^R (bacterial)	this work

Construction of complementing vectors

pCMBL2

To generate the basic binary vector suitable for complementation experiments in the used transgene system in *A. thaliana*, binary vector *pCMBAR* (pCAMBIA-proMAS-BAR-35ster) was used as a starting point (M.F. Mette, personal communication). A PCR product containing *LacZ* and MCS of plasmid *pGEM-7Zf(-)* was amplified using primers pGEM7Z-MCS-for and pGEM7Z-MCS-rev (Table 3) and introduced into the unique PmeI restriction site of vector *pCMBAR* to obtain vector *pCMBL2*. *pCMBL2* contains unique AatII, ZraI, PspXI, SciI, XhoI, XmaI, SmaI, BstBI, HindIII, BspEI and BstXI restrictions sites suitable for the insertion of DNA fragments with the option for using X-Gal-based blue / white screening of bacterial colonies containing recombinant plasmids.

pCMBL2+ProIDN2:IDN2

A PCR product comprising the wild-type *IDN2* ORF (including 3'UTR) and a region of around 1300 bp upstream of the transcriptional start site was amplified from *A. thaliana* accession Col-0 genomic DNA. Amplification was carried out using 0.5 U PhusionHF polymerase (Thermo Fischer, Langenselbold) in 50 µl volume containing 1x HF Buffer, 0.25 mM dNTPs, 0.75 µM primers *IDN2-clone-for* and *IDN2-clone-rev* (Table 6) and 2 µl of genomic DNA preparation. The primers used introduced XhoI restriction sites at the ends of the PCR product that were suitable for later sub-cloning into *pCMBL2*. PCR products were separated by electrophoresis in 0.8% agarose gels. The band corresponding to the *IDN2* PCR product was cut from the gel and the contained DNA was purified using a QIAquick gel extraction kit (Qiagen, Hilden) and cloned into vector *pSC-A-amp/kan* using a Strataclone PCR cloning kit (Agilent Technologies). The inserts of ß-galactosidase-negative clones were sequenced using standard primers M13-for and *M13-rev* as well as *IDN2*-specific primers (Table S2). One clone containing the unaltered wild type sequence was further propagated. The *IDN2* gene was excised by XhoI cleavage and cloned into the XhoI restriction site of *pCMBL2* yielding the binary vector *pCMBL2+ProIDN2:IDN2* suitable for complementation tests by *Agrobacterium tumefaciens* (*A. tumefaciens*)-mediated transformation.

pCMBL2+ProDRM2:DRM2

A PCR product comprising the wild-type *DRM2* ORF (including 3'UTR) and a region of around 450 bp upstream of the transcriptional start site was amplified from *A. thaliana* accession Col-0 genomic DNA. Amplification was carried out using 0.5 U PhusionHF polymerase (Thermo Fischer, Langenselbold) and primers DRM2-AatII-F and DRM2-BspEI-R (Table 6) in 50 µl volume containing 1x HF Buffer, 0.25 mM dNTPS, 0.75 µM primers DRM2-AatII-F and DRM2-BspEI-R and 2 µl of genomic DNA to obtain a PCR product flanked by AatII and BspEI restriction sites. The PCR product was purified by agarose electrophoresis, cut from the gel, extracted using QIAquick gel extraction kit (Qiagen, Hilden) and cloned into vector *pSC-A*-amp/kanusing a Strataclone PCR cloning kit (Agilent Technologies). The inserts of ß-galactosidase-negative clones were sequenced using standard primers *M13-for* and *M13–rev* as well as *DRM2*-specific primers (Table S2). One clone containing the correct wild type *DRM2* ORF was further propagated. The plasmid containing the functional *DRM2* gene as well as vector *pCMBL2* were cleaved using AatII and BspEI. The excised functional *DRM2* gene was

then cloned into vector *pCMBL2* yielding the vector *pCMBL2+ProDRM2:DRM2* suitable for complementation tests by *A. tumefaciens*-mediated transformation.

Table 6: Primers used for cloning procedures

Product	Primer	Sequence (5' ➔ 3')	Restrcition Site
MCS of pGEM7Zf(+)	*pGEM7Z-MCS-for*	CGCGTTTAAACAGGGCGCGTCCATTCGCCATTC	PmeI
	pGEM7Z-MCS-rev	CATGTTTAAACGGAAGAGCGCCCAATACGC	PmeI
ProIDN2:IDN2	*IDN2-clone-for*	CTTGACTCGAGACTTGCCTTGTGTCAGCG	XhoI
	IDN2-clone-rev	ACGCTCGAGGGGTCAATATCAAATTTGAC	XhoI
ProDRM2:DRM2	*DRM2-AatII-F*	GTATGTGACGTCCTTTGACTCGCCGGTCGCTAC	AatII
	DRM2-BspEI-R	GACTCCGGAACCAAAGTGTCTATACAAATAC	BspEI

2.5 *E.COLI* CULTURE AND PLASMID PREPARATION

For bisulfite sequencing analysis 1.5 ml of LB medium supplied with 50 mg/l ampicillin or 50 mg/l kanamycin was inoculated with a single ß-galactosidase-negative colony and cultivated over night at 37 °C and 200 rpm shaking. Plasmid was isolated using QiaPrep Spin Mini Kit (Qiagen, Hilden) according to the manufacturer's instructions.

For propagation of complementing binary vectors, 5 ml of LB medium supplied with 50 mg/l kanamycin was inoculated with a single colony and cultivated for 6 h at 37 °C and 200 rpm shaking. Subsequently 50 ml of LB medium supplied with kanamycin were inoculated with 2 ml of this pre-culture and further incubated over night at 37 °C and 200 rpm shaking. Plasmid was isolated using Qiagen Plasmid Midi Kit (Qiagen, Hilden) according to manufacturer's protocol.

2.6 STABLE TRANSFORMATION OF *A. THALIANA* PLANTS

Stable transformation of *A. thaliana* was carried out by *A. tumefaciens*-mediated gene transfer *via* floral dip transformation using *A. tumefaciens* strain pGV2260 (Deblaere *et al.,* 1985; Clough and Bend 1998).

Transformation of *A. tumefaciens* by electroporation

Electrocompetent *A. tumefaciens* cells were obtained using a modified protocol according to Mersereau *et al.,* (1990). LB medium supplied with 20 mg/l rifampicin was inoculated with a single colony of *A. tumefaciens* strain pGV2260 and incubated under shaking at 28 °C till the culture reached an OD_{600} of about 0.5. Bacteria were sedimented by centrifugation for 20 min at 5600 x g at 4 °C. The resulting pellet was washed twice and finally resuspended in 10 ml of pre-chilled 10% (v/v) glycerin. Alliquotes of 50 µl of

transformation competent *A. tumefaciens* strain pGV2260 were stored at -80°C until further use.

Prior to transformation into *A. tumefaciens*, DNA was desalted by microdialysis. For dialysis, the solution containing plasmid DNA was spotted on a MF-Millipore membrane (Millipore, Billerica, USA; pore size 0.025 µm) floating on bi-distilled water and incubated for 15 min at ambient temperature. For transformation, 50 µl of electrocompetent *A. tumefaciens* cells were inoculated with 10 µg of desalted plasmid DNA. The suspension was transferred to a pre-chilled electroporation cuvette and two pulses of 2300 mA were applied using a Gene PulserII device (Bio-Rad Laboratories GmbH, Munich). Transfected cells were suspended in 800 µl SOC medium and incubated at 28°C for 1 h. 100 µl of preculture were spread on solid LB medium supplied with 20 mg/l rifampicin and 50 mg/l kanamycin and incubated at 28°C for 48 to 72 h. Appearing colonies were streaked out on solid LB medium containing 50 mg/l kanamycin and in parallel used to inoculate 1.5 ml of LB medium supplied with 50 mg/l kanamycin. Plates and cultures were incubated for 16-20 h at 28°C. A QIAprep Spin Mini Kit (Qiagen, Hilden) was used for plasmid isolation from 10 ml of liquid culture according to the manufacturer´s protocol. Presence of the correct insert in the contained plasmid was checked by insert-specific PCR using 1 µl of plasmid solution as template followed by PCR product analysis *via* agarose gel electrophoresis. If products of expected size were detected, the respective *A. tumefaciens* strain was considered suitable for transformation of *A. thaliana*.

Transformation of *A. thaliana*

A. thaliana plants designated for transformation were grown for approximately 4 weeks at long day conditions until onset of flowering.

A single *A. tumefaciens* colony positive for the plasmid was used to inoculate 30 ml of YEB medium supplied with 20 mg/l rifampicin and 50 mg/l kanamycin and incubated in an orbital shaker at 200 rpm and 28°C until an OD_{600} of approximately 0.8. 10 ml each of this culture were used to inoculate two times 200 ml of YEB medium supplied with 50 mg/l kanamycin. After incubation for another 24 h at 200 rpm and 28°C, the bacteria were spinned-down for 5 min at 5000 rpm in a Sorvall RC5B centrifuge using rotor Sorvall SH4. Bacteria collected from 400 ml culture were re-suspended in 800 ml infiltration medium containing 0.1 mg/ml acetosyringone. After addition of Silvet L-77 to a final concentration 0.025% (v/v) (Leehle Seeds, Roundrock, USA), stalks of flowering *A. thaliana* plants were submerged in the *A. tumefaciens* suspension. Infiltrated plants

were kept in darkness for 16 h and subsequently cultivated under a long day regime until seed ripening.

To identify primary transformants, T_1 generation plants were screened for BASTA resistance (BASTA^R). Approximately 1.5×10^4 seeds from transformation were germinated on soil under a long day regime. After appearance of the second pair of true leafs, plants were sprayed with 300µM BASTA solution. Spraying was repeated after additional five days of cultivation. BASTA^R plants were transferred to small pots and further cultivated under a long day regime until seed ripening. Seeds were harvested from individual plants.

2.7 SURFACE STERILIZATION OF *A. THALIANA* SEEDS

For surface sterilization, *A. thaliana* seeds were incubated for 5 min in 70% ethanol and for 10 min in 8% NaOCl solution. Subsequently, seeds were washed four times in sterile distilled H_2O and resuspended in 0.1% sterile agarose in water.

2.8 DNA EXTRACTION FROM *A. THALIANA* LEAF TISSUE

For 50 to 100 mg fresh weight of leaf tissue, DNA isolation was performed using DNeasy Plant Mini Kit (Qiagen, Hilden, GER), whereas isolation from 500 mg to 1 g was performed using DNeasy Plant Maxi Kit (Qiagen, Hilden, GER) according to the manufacturer´s protocol.

2.9 RNA ISOLATION PROCEDURE

For isolation of preparations enriched in "long" and "small" RNAs from of *A. thaliana* leafs, a two-step procedure employing columns of the Qiagen RNeasy Maxi and Qiagen RNeasy Midi Kits were used according to a modified version of the protocol "Purification of miRNA from animal and plant tissues using RNeasy Lipid Tissue Kit and RNeasy MinElute Cleanup Kit" provided by the manufacturer. Plants were grown for 8 weeks under a short day regime.

Isolation of "long" RNAs and reverse transcription-PCR

For isolation of "long" RNAs, approximately 500 mg of leaf material were flash frozen in liquid nitrogen and ground by vortexing for one minute after addition of five grinding spheres (Roche Diagnostics, Mannheim). The material was re-suspended in 15 ml of TRIzol reagent (Invitrogen GmbH, Karlsruhe) or QIAzol reagent (Qiagen, Hilden) and mixed by vortexing at room temperature for one minute. The resulting suspension was

transferred into Corex 30 ml glass centrifugation tubes (Thermo Fisher Scientific, Langenselbold). After adding 3 ml of chloroform, the tubes were sealed with Parafilm M (Brand GmbH, Wertheim), vortexed for one minute at ambient temperature and centrifuged for 30 minutes at 8140 x g at 4°C in a Sorvall RC5B Plus centrifuge (Thermo Fisher Scientific, Langenselbold). The upper, polar phase (V_1) was transferred to a 50 ml centrifugation tube and mixed with 1 time V_1 of 80% ethanol. The mixture was transferred to an RNeasy Maxi Kit column and centrifuged for five minutes at 2780 x g at 20°C. The flow-trough (V_2) was saved for "small" RNA preparation and stored on ice (see below), while the column with bound "long" RNA was washed once with 15 ml of buffer RW1 and twice with 15 ml of buffer RPE. Wash buffer was removed by 5 min of centrifugation at 2780 x g and room temperature. RNA was eluted by incubation in 1.2 ml of RNAse free water and centrifugation at 2780 x g for 5 min. The RNA concentration was determined spectrophotometrically using a Nanodrop ND-1000 (PeqLab, Erlangen).

Isolation of "small" RNAs

The flow-through of the RNeasy Maxi Kit column (V_2, see above) was transferred to a new vessel, mixed with 1.4 times V_2 of 100% ethanol and applied to an RNeasy Midi Kit column by consecutive centrifugations for 5 minutes at 2780 x g at 20°C. The column with bound "small" RNA was washed twice with RPE buffer. Remaining buffer was removed by an additional centrifugation step. Subsequently, 250 µl of RNase free water (V_E) were added and the column was centrifuged for 5 min at 2780 x g at 20°C. After elution, the RNA concentration was determined using an Ultrospec 3100pro UV/Vis spectrophotometer.

The eluted "small" RNA was precipitated by addition of 0.11 times V_E of 3 M sodium acetate (pH 5.2) and 2.5 times V_E 100% ethanol and subsequent incubation at -20°C for at least 12 hours. Precipitated RNA was collected by centrifugation for 5 minutes at 18000 x g at 4°C. The supernatant was removed and the sedimented RNA was washed once in 70% ethanol and dried under vacuum at room temperature for 30 minutes. Finally, the isolated "small" RNA was dissolved in 25 µl of RNase-free water and stored at -20°C until further use.

2.10 REVERSE TRANSCRIPTION PCR

For cDNA synthesis, possibly contaminating genomic DNA was fragmented with DNAseI (Fermentas, St. Leon - Rot) at 37°C for 30 min. The reaction was stopped by adding 1 µl 25 mM EDTA and incubation for 10 min at 65°C. 1 µg of RNA was reverse-transcribed

using a RevertAid H Minus first strand cDNA synthesis kit (Fermentas, St. Leon - Rot). 1 µl of the reverse transcription reaction were used as template for subsequent amplification by PCR using specific primer pairs (Table 7).

Table 7: Primers used for RT-PCR

Gene	Primer	Sequence (5' ➜ 3')	Product (bp)	Reference
IDN2	*qIDN2-for*	TCAGATTGGCATCCATTCAA	216	Finke *et al.,* 2012b
	qIDN2-rev	CCGCATTGTAAGGACCATCT		
Actin2	*Act2-for*	GGTTGTGTCAAGAAGTCTTGTGTACTTTAGTTTTA	245	Johnson *et al.,* 2002
	Act2-rev	ATAGCTGCATTGTCACCCGA		

2.11 QUANTIFICATION OF *PRONOS* SIRNAS

Small RNA northern Blots

For denaturing polyacrylamide gel-based separation of "small RNA" (Mette *et al.,* 2005), 25 µl of RNA gel loading buffer II (Life Technolgies GmbH, Darmstadt) were added to the 25 µl of "small" RNA preparation. In addition, NEB microRNA marker (NEB, Frankfurt am Main) containing RNAs of 17 nt, 21 nt and 25 nt length was included as size standard. The mixtures were incubated at 95°C for 5 min and then submitted to electrophoretic separation (800 V; maximum 10 W) for approximately 60 min on 15% polyacrylamid gels containing 7 M urea. After electrophoresis, the gels were incubated in 0.5 x TBE buffer supplied with 1 µg/ml ethidium bromide for 10 minutes and then washed in 0.5 x TBE without ethidium bromide for 20 min. Equal loading of lanes with RNA was checked under UV illumination.

Separated RNAs were transferred to Zeta-Probe GT nylon membranes (Bio-Rad Laboratories, Munich) using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad Laboratories, Munich). From anode to cathode, 3 sheets of 3 MM GB003 blotting paper soaked with 0.5 x TBE buffer, a 0.5 x TBE buffer-wetted sheet of cut-to-size nylon membrane, the polyacrylamide gel and again 3 sheets of 3MM GB003 blotting paper soaked with 0.5 x TBE buffer were set up. Transfer was carried out at 10 V for 1 h in a cold lab. Subsequently, the membrane was rinsed in 0.5 x TBE and the transferred RNA was crosslinked to the membrane by incubation in vacuum at 80°C for 2 h.

Preparation of template DNA for probe synthesis by *in vitro* transcription

Plasmid DNA containing the *ProNOS* template sequence downstream of a T7 promoter was cleaved overnight in a 200 µl reaction with restriction enzyme BamHI (Fermentas, St. Leon - Rot) at 37°C. After restriction cleavage, the plasmid DNA was purified by

phenol-chloroform extraction. 300 µl of Tris/HCl saturated phenol-chloroform-isoamylalcohol were added and mixed with the restriction setup by vortexing for 1 min. Phase separation was achieved by centrifugation for 5 minutes at 18600 x g and 8°C. Phenol-chloroform extraction was repeated once. Subsequently, 300 µl of chloroform were added, mixed with the aqueous phase by vortexing for 1 min and centrifuged for 5 min at 18600 x g at 8°C. The aqueous phase was transferred to a new reaction tube. Plasmid DNA was concentrated by ethanol precipitation, dried under vacuum and dissolved in 50 µl of RNase-free water.

Radioactive labeling of RNA probes.

The *ProNOS* sense siRNA-specific probe was labeled by *in vitro* transcription. 1 µg of template DNA, 2 µl 10x transcription buffer, 1.5 µl of *rNTP*s (rATP, rCTP, rGTP) (6.7 mM each), 1 µl 0.2 M DTT, 1 µl Ribolock RNase inhibitor (Fermentas, St. Leon -Rot), 1 µl T7 polymerase (Fermentas, St. Leon -Rot) and 12.5 µl of [α-^{32}P] UTP (Hartmann Analytic, Braunschweig) were combined and incubated for 2 h at 37°C. The DNA template was fragmented by addition of 10 U of RNase-free DNAse I and further incubation at 37°C for 15 min. After addition of 300 µl Na_2CO_3/$NHCO_3$ (120 mM/80 mM) and incubation for 2.5 h at 60°C for random RNA fragmentation, 20 µl of 3M NaOAc/HOAc (pH 5.0) were added for neutralization.

Radioactive labeling of mir167 probe

Radioactive labeling of the miR167 DNA oligonucleotide (5'–TAGATCATGCTGGCAGCTTCA-3') probe (Wu *et al.,* 2006) was carried out using T4 Polynucleotide Kinase (Fermentas, St. Leon - Rot) according to manufacturer´s protocol.

Hybridzation and detection of small RNAs

For detection of *ProNOS* sense siRNA, the membrane with transferred "small" RNA was pre-hybridized with 40 ml hybridization solution at 42°C for at least two h. After pre-hybridization, the hybridization solution was replaced and 300 µl of freshly prepared ^{32}P-labled RNA probe was added. Hybridization was performed at 42°C overnight and the membrane was washed two times (15 min each) in buffer containing 2 x SSC/0.2% SDS at room temperature. The membrane was wrapped in plastic foil and exposed to X-ray film with intensifier screen for 1, 4 and 7 days at -80°C.

For re-hybridization of membranes with a miR167 specific probe, the hybridized *ProNOS*-specific probe was stripped of by brief washes in buffer containing 0.1 x

SSC/0.5% SDS at 95°C. The stripped membranes were pre-hybridized with hybridization buffer according to Church and Gilbert for 16 h at 42°C. Subsequently, the labeled DNA oligonucleotide probe was added and allowed to hybridize for 24 h at 42°C. The blot was washed in 2 x SSC containing 0.2% SDS and exposed to X-ray film with intensifier screen for 3 days at -80°C.

2.12 DNA METHYLATION ANALYSIS

Analysis by methylation-sensitive restriction cleavage

For *TARGET-ProNOS* methylation analysis, approximately 50 ng of DNA extracted from adult leaves of 8-week-old plants grown under short day regime were dissolved in 400 µl of distilled water, 50 µl of 10x Tango buffer and 50 µl of bi-distilled water were added to a final volume of 500 µl. Aliquots of 100 µl were incubated with 10 U of restriction enzymes Psp1406I, NheI, Alw26I, NcoI (Fermentas, St. Leon –Rot) or without restriction enzyme, respectively, and incubated at 37°C for 16 h. Subsequently, the restriction enzymes were heat-inactivated by incubation for five minutes at 85°C. 399 µl of bi-distilled water were added to a final volume of 500 µl. Quantitative PCR was performed using an iCycler IQ device (Bio-Rad Laboratories GmbH, Munich). 12.5 µl of SYBR Green Supermix (Bio-Rad Laboratories GmbH, Munich) and 1.25 µl of primers *ProNOS-top-F* and *ProNOS-top-R* (final concentration 0.25 µM each) were added to 10 µl of cleaved DNA or control DNA templates, respectively, to reach a final volume of 25 µl for each sample. The PCR was performed using the following temperature regime:

Temperature	Duration	Cycles
95°C	5 min	1
95°C	15 sec	
62°C	30 sec	40
72°C	30 sec	
	Data acquisition	
72°C	5 min	1

PCR was calibrated using logarithmic serial dilutions from 10^{-2} to 10^{-5} of genomic DNA preparations. The threshold cycle (C_t value) for reactions with serial dilution samples was determined in technical duplicates. C_t values of reactions with cleaved and control DNA samples were determined in duplicate. Samples with mean C_t values ± 0.25 cycles were included in calculations. Data analysis was performed using the ΔΔCt method (Pfaffl, 2001). Results are presented as percent of the mean signal obtained for the control samples without restriction enzyme (set to 100%).

For methylation analysis of *IGN5A*, *IGN23* and *IGN25* sequences, 0.05 µg of genomic DNA was dissolved in 25 µl bi-distilled water and 5 µl of Buffer R and 20 µl of bi-distilled water were added to a final volume of 50 µl. After addition of 10 U of restriction enzyme BsuRI/HaeIII (Thermo Fischer, Langenselbold), the reaction was incubated for 16 h at 37°C. Subsequently, the reaction was incubated for 5 min at 85°C for restriction enzyme inactivation.

Table 8: Primers used for methylation analysis by methylation sensitive restriction cleavage

Target	Name	Sequence (5' → 3')	Reference
ProNOS	*ProNOS-top-F*	GATAGTTGGCGAAATTTTCAAAGTCC	Finke *et al.,* 2012b
	ProNOS-top-R	TGCAATCCATCTTGTTCAACCATGG	
IGN5	*IGN5A-F*	TCCCGAGAAGAGTAGAACAAATGCTAAAA	Wierzbicki *et al.,* 2008
	IGN5A-R	CTGAGGTATTCCATAGCCCCTGATCC	
IGN22	*IGN22-F*	CAAAAATATTCACCCGCTACAAACAAAAA	Rowley *et al.,* 2011
	IGN22-R	TCTTCCATTTGTGGGGCATGGT	
IGN23	*IGN23-F*	ACTGAAAATTGTAAACAAAGAAACGGCACTACA	Rowley *et al.,* 2011
	IGN23-R	GATCGGTCCATAAACTTGTTGGGTTT	
IGN25	*IGN25-F*	CTTCTTATCGTGTTACATTGAGAACTCTTTCC	Rowley *et al.,* 2011
	IGN25-R	ATTCGTGTGGGCTTGGCCTCTT	

PCR was performed in a reaction volume of 25 µl in an Eppendorf MasterCycler Gradient device (Eppendorf, Hamburg) using GoTaq Flexi Polymerase (Promega, Mannheim). 14.4 µl bi-distilled water, 5 µl of GoTaq Flexi Polymerase Buffer, 3 µl of 25 mM MgCl$_2$ (final concentration: 3 mM), 0.5 µl 10 mM dNTPs (final concentration: 0.2 mM), 0.5 µl of template specific forward and reverse primer (Table 7) (final concentration 0.2 µM each) and 1.25 U GoTaq Flexi polymerase were combined and added to 1 µl of restricted template DNA. PCR reaction was performed using the following temperature regime: 5 min 95°C, 35x (15 sec 95°C, 10 sec 61°C, 20 sec 72°C). Amplification of ING22, a sequence lacking a HaeIII cleavage site, was used as loading control.

Bisulfite sequencing

For methylation analysis by bisulfite-mediated conversion of unmethylated cytosines into uracil, approximately 0.15 µg of genomic DNA extracted from leaves of 8-week-old plants grown under short day conditions were bisulfite-treated using an Epitect Bisulfite Kit (Qiagen, Hilden) following the protocol "Sodium Bisulfite conversion of unmethylated cytosines in DNA from Low-concentration solution (Version 2009)" provided by the manufacturer.

For PCR amplification of target sequences, GoTaq Flexi DNA polymerase (Promega, Mannheim) was used. Reaction was carried out in a volume of 50 µl containing 10 µl of 5x GoTaqGreen Flexi buffer, supplied with 3 mM MgCl$_2$, 0.2 mM dNTPs, 0.2 µM of respective forward and reverse primers and 1 U of GoTaq polymerase. 1 µl of converted DNA was used as template for amplification. Primers used for amplification of *TARGET ProNOS*, *AtSN1, MEA-ISR, AtMU1, AtCOPIA4. IGN5A, IGN23* and *BASHO210* regions are listed in Table 9.

Table 9: Primers used for bisulfite sequencing

Target	Name	Sequence (5' ➔ 3')	Reference
ProNOS	*pNOS bitop2f*	AATTTGTTGGTTATTATATGATAGTTG	this thesis
	pNOS bitop3r	AACCTACATACAATCCATCTTATT	
AtSN1	*AtSN1-5F*	GTTGTATAAGTTTAGTTTTAATTTTAYGGATYAGTATTAATTT	Zhen *et al.,* 2007
	AtSN1-3R	CAATATACRATCCAAAAAACARTTATTAAAATAATATCTTAA	
MEA-ISR	*JP1026*	AAAGTGGTTGTAGTTTATGAAAGGTTTTAT	Cao and Jacobsen,2002
	JP1027	CTTAAAAAATTTTCAACTCATTTTTAAAAAA	
AtMU1	*JP1387*	ATATCCTTCTCTTTCATTCARATTTTAATTTTTTCCRT	Bäurle *et al.,* 2007
	JP1388	GTTTAGTGTTTATGATTATATAATTGTGTTATAATTGTTAAT	
AtCOPIA4	*JP3100*	GGTTGTYTGTGTTTTTTATGGTTYAGATTTTATA	Johnson *et al.,* 2007
	JP3101	ATAACTRAACCACARATTCARACCCATTTTCATTT	
IGN5	*BS-IGN5A-F*	YAYATTTGTTTAGGAAATATGTTAGTAAG	this thesis
	BS-IGN5A-R	ATCATRCATTCTATTRCCCATARCA	
IGN23	*BS-IGN23-F*	TAGATTTGGTTAGGTAAAGGTTAG	this thesis
	BS-IGN23-R	TTTTTTAAATATAAAAACTAARCCTAC	
BASHO210	*BS-5-B210-F*	GGAGATGTATATAAYTAGTAGG	this thesis
	BS-5-B210-R	ATACCCRATCCRAATACCCAAAC	

PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Hilden), cloned into vector pSC-A and transferred into Strataclone Solo Pack competent cell using the Strataclone PCR cloning kit (Agilent Technologies, Santa Clara, USA) according to manufacturer´s protocol. After recovery, transformed cells were spread on LB plates containing either 50 mg/l ampicillin or 50 mg/l kanamycin, IPTG and X-Gal and incubated at 37℃ for 16 h. 12 to 24 colonies negative for ß-galactosidase activity were used to inoculate 1.5 ml LB medium cultures, which were then incubated for 12 to16 h at 37℃ in an orbital shaker at 200 rpm. Subsequently, a QIAquick SpinMini Kit was used to isolate plasmid DNA.

For validation of the correct insert size, 10 µl of plasmid preparation was restriction-cleaved using 4 U EcoRI and 4 µl Buffer 10x Tango in 20 µl final volume. The reactions were incubated for 3 h at 37℃. Restriction fragments were separated by agarose gel electrophoresis and detected using UV excitation in presence of 1 µg/l ethidium bromide.

Uncleaved plasmid DNA from preparations that had shown a restriction fragment of the expected PCR product size ± 30 bp were Sanger-sequenced by IPK´s in-house core facility service.

2.13 PROTEIN QUANTIFICATION

Total protein in the same extracts was determined using a Pierce BCA Protein Assay kit (ThermoFisher, Rockford, USA). 25 µl of the NPTII ELISA raw extract were added to 500 µl of BCA working solution and incubated for 30 min at 37°C in a water bath. After incubation, 500 µl of bi-distilled water were added to every sample and extinction at 592 nm was determined using an Ultrospec 3100pro UV/Vis spectrophotometer (Amersham Bioscience, Freiburg, GER). Extinction values were converted to protein concentration values using a BSA serial dilution in concentration range between 0.125 and 2 µg/ml as reference.

Amounts of NPTII protein in plant extracts were determined using Agdia PathoScreen Kit for NPTII (Agdia, Elkhart, USA). Rosette leafs of 8-week-old short-day-grown plants were flash frozen in liquid nitrogen, ground using a swing mill type Retsch MM301 and resuspended in provided protein extraction buffer. All further procedures were performed according to manufacturer's recommendations.

2.14 DEVELOPMENT OF INDEL AND CAPS MARKERS

InDel markers

Initially, PCR-based markers detecting insertion / deletion (InDel) sequence polymorphisms between *A. thaliana* accessions Col-0 and L*er* were used to map the positions of mutations induced by EMS. Per chromosome, four InDel markers evenly distributed along the respective chromosome arms were chosen from a set published by Salathia *et al.,* (2008). Selection was according the following criteria:

- Maximal PCR product size: 1000 bp
- Minimal size difference of products from Col-0 compared to L*er*: 25 bp
- Maximal size difference of products from Col-0 compared to L*er*: 500 bp

Prior to application in mapping procedure, InDel markers were tested with DNA from Col-0 and L*er* plants for reliability. The PCR primers used and the sizes of resulting PCR products are indicated in Table S3.

CAPS marker

Cleaved amplified polymorphisms (CAPS)-based markers were developed based on single nucleotide polymorphisms (SNPs) between *A. thaliana* accessions Col-0 and L*er* annotated in the TAIR database (versions TAIR 8 to10). Col-0 and L*er* sequence were checked for suitable restriction sites using software webcutter 2.0 (rna.lundberg.gu.se/cutter2/).

Prior to application in mapping procedure, CAPS markers were tested with DNA from Col-0 and L*er* plants for reliability. The PCR primers, the appropriate restriction enzymes, the sizes of resulting PCR products and restriction fragments are indicated in Table S5.

2.15 ILLUMINA VERACODE GOLDEN GATE ASSAY

Development of the 48-plex assay

For the Illumina Veracode GoldenGate assay (GoldenGate assay) 47 SNPs variant between accessions Col-0 and L*er*, were chosen based on a set of published SNP marker suitable for multiplexed genotyping (Kover *et al.,* 2010). Markers were chosen according to criteria provided by Illumina Inc. (www.illumina.com/documents/products/technotes/technote_goldengate_design.pdf).

At first, 60 bp of flanking sequence on either side of chosen SNPs were checked for additional variant positions using sequence information of two publicly available databases (www.arabidopsis.org TAIR 9.0 and www.signal.salk.edu/atg1001.htm). SNPs with flanking sequences that contained additional polymorphisms in this range according to one of these databases were excluded from further consideration. As second selection criterion, the threshold for the "minimum SNP specific score" as calculated by Illumina Inc. according to a confidential algorithm, was set to 0.75 in order to achieve at least three evenly distributed reliable markers per chromosome arm. In addition, one marker specifically targeting the sequence of the *H* transgene was included in the assay. All SNPs included in the GoldenGate assay are listed in Table S4. Tests were performed according to the protocol "GoldenGate genotyping assay for VeraCode – Manual Protocol Rev. B" provided by the manufacturer, without performance of the optional steps "Make qDNA plate" and "Scan qDNA plate". The obtained data were analyzed using the GenomeStudio software provided by Illumina Inc. according to provided guidelines. Only markers matching a GT-Score of at least 0.6 in the assay output were included in the genotyping analysis.

2.16 NEXT GENERATION SEQUENCING

Next generation sequencing (NGS) of plant genomic DNA was performed by IPK's in-house sequencing facilities using an Illumina HiSeq2000 device according to manufacturer's protocols. Sequencing was performed in paired-end reads. Mapping of obtained reads to the *A. thaliana* reference genome (TAIR10), SNP calling, effect prediction and *de novo* assembly of the *K/K*; *H/H* and mutant genomes was performed by Dipl. math. Martin Mascher, using the Burrow-Wheeler Alignment Tool (http://bio-bwa.sourceforge.net, Li and Durbin, 2009) the Sequence Alignment/Map tool (http://samtools.sourceforge.net/, Li *et al.,* 2009), snpEff (http://snpeff.sourceforge.net/, Cingolani *et al.,* 2012) and CLC assembler (www.clcbio.com), respectively according to established protocols. For identification of mutations in the candidate lines, the respective SNP lists were imported into MS Excel and compared using the Merge Table Wizzard Add-in (http://www.ablebits.com/excel-lookup-tables/index.php). SNPs occurring in more than one of the mutant candidate lines were considered as preexisting in the transgenic line submitted to mutagenesis and thus considered as "false positives" and excluded from further analysis. Furthermore, "true positive" SNPs affecting promoters, introns, 5' UTR and 3'UTR as well as such in exons that were expected to cause silent mutations were excluded from searches for mutations plausibly causative for kanamycin resistance.

2.17 MUTAGENESIS

Custom mutagenesis of a line homozygous for *TARGET* transgene $K_{chr1-10}$ (briefly *K*) and *SILENCER* transgene (briefly *H)* (Fischer *et al.,* 2008) was performed by Leehle Seeds, Round Rock, Texas, USA. Approx. 100,000 seeds of the F_4 generation were exposed to ethyl methanesulfonate (EMS) and then allowed to germinate on soil in order to obtain approx. 50,000 mutagenized M_1 plants, which were allowed to set seeds by self-pollination. The resulting M_2 seeds were harvested in 32 batches, each batch thus containing the progeny of approximately 1,500 M_1 plants. A sample of the initial F_4 seeds was germinated without EMS treatment under otherwise same growth conditions and the resulting non-mutagenized control (C_1) plants were also allowed to self-pollinate to generate C_2 seeds.

2.18 ESTABLISHMENT OF MAPPING POPULATIONS

To establish mapping populations, M_3 plants of mutants were crossed with the *A. thaliana* accession L*er*. The success of the crosses was validated in F_1 progeny (M_3F_1)

by qualitative histochemical GUS staining to detect a *Pro35S-GUS* reporter gene present in the *K* transgene. In order to insure the presence of the *K* and *H* transgenes as well as of the mutation in homozygous state, F_2 generation seedlings obtained from these crosses by self-pollination (M_3F_2) were screened for plants showing hygromycin and kanamycin resistance (Hyg^R Kan^R) simultaneously. To rule out possible "false positive" M_3F_2 individuals, approx. 110 plants of M_3F_3 progeny obtained by self-pollination was germinated on GM supplied with kanamycin (200 mg/l) and hygromycin (20 mg/ml). Only M_3F_2 plants that met the expected minimal 56% of Hyg^R Kan^R plants in their respective F_3 progeny were included in the mapping populations. In addition, C_3 non-mutagenized control plants were crossed to L*er* and resulting C_3F_1 progeny was allowed to self-pollinate. Resulting C_3F_2 generation seeds were screened for Hyg^R plants by germination on GM supplied with 20 mg/l hygromycin. Presence of *H* was validated by qualitative histochemical GUS staining. Resulting C_3F_2 Hyg^R GUS^+ plants containing *K* and *H* transgenes were used as control mapping population.

Histochemical GUS stain

The ß-glucuronidase (GUS) histochemical staining procedure was carried out in 96-well micro titer plates (Greiner Bio-One, Solingen, GER). Two leaf discs of distinct mature rosette leafs were assayed per plant. 200 µl of GUS working solution were added to each well and allowed to infiltrate the leaf disc by application of vacuum for 5 minutes. After incubation for 16 h at 37°C, the GUS working solution was replaced by 200µl of 70% ethanol and incubation for additional 3 h was allowed. This ethanol washing step was repeated until all chlorophyll was removed from leave disks.

Table 10: Solutions used for histochemical GUS staining

Solution	Component	per 10 ml	final conc.
X-Glu solution	X-Glu (5-bromo-4-chloro-3-indolyl-ß-glucuronide) in DMFA (N,N-dimethylformamid)	1 g	1% (w/v)
NaN$_3$ solution	NaN$_3$ in sterile dest. H$_2$O	0,5 g	5% (w/v)
working solution	X-Glu solution NaN$_3$ solution 100 mM NaH$_2$PO$_4$ / Na$_2$HPO$_4$ pH 7.0	0,5 ml 0,2 ml 9,3 ml	0,05 % (w/v) 0,1 % (w/v) 93 mM

2.19 STATISTIC ANALYSIS

To test DNA methylation levels of identified mutant lines determined by bisulfite sequencing for significant differences relative to *K/K;H/H* DNA methylation X^2 test was used. X^2 test was also used for statistic analysis of segregation of the homozygous Col-0 allele in mapping populations.

3 RESULTS

3.1 FORWARD GENETIC SCREEN FOR MUTATIONS RELEASING RNA-DIRECTED TRANSCRIPTIONAL GENE SILENCING

Selection of mutants using kanamycin resistance as indicator

In order to identify mutants in which RNA-directed transcriptional gene silencing (RdTGS) of the *ProNOS-NPTII* reporter gene was released, and thus *NPTII* expression was reactivated, approx. 20,000 M_2 seeds per obtained seed batch (see section 2.17) were germinated on growth medium (GM) containing 200 mg/l of kanamycin. All M_2 batches yielded some kanamycin resistant (Kan^R) individuals, while no Kan^R plant was observed in the non-mutagenized C_2 control. From each batch, ten Kan^R M_2 plants (*1-1* to *1-10*, *2-1* to *2-10*, and so on, till *32-1* to *32-10*) were transferred to soil and allowed to self-pollinate. Resulting M_3 seeds from individual M_2 were collected and their kanamycin resistance was verified by germinating approx. 200 seeds per M_2 line on GM containing 200 mg/l of kanamycin. This resulted in the identification of 104 Kan^R M_3 lines that showed at least 95% of viable resistant seedlings in the presence of kanamycin. To avoid possible redundant siblings, initially one Kan^R M_3 line per M_2 batch was chosen for characterization.

As the first step of my thesis work, the presence and integrity of the *K* and the *H* transgene in Kan^R M_3 lines was checked by PCR using transgene-specific primer combinations (Table S1). It revealed that all 32 initially selected Kan^R lines contained the *K* transgene, but only a single line contained the *H* transgene (Figure 3B and data not shown). As loss of the *H* transgene *per se* can result in the *ProNOS-NPTII* reactivation (Aufsatz *et al.,* 2002), the presence of both transgenes is a *de regieur* prerequisite for the identification and later map-based cloning of gene loci essential for RdTGS. Thus, so far only Kan^R M_3 line *2-5* was suitable for further work.

To obtain additional candidates suitable for further analysis and map-based cloning, the remaining 72 Kan^R M_3 lines from the first screening were tested for the presence of full length *K* and *H* transgenes using transgene-specific PCR. Five additional lines were found to contain both transgenes and were included in further analysis (Figure 3C).

Selection of mutants using combined hygromycin and kanamycin resistance as indicator

As the above six confirmed Kan^R M$_3$ lines were rather limited material to work with, it was attempted to isolate further candidate lines by repeating the screening of the M$_2$ material.
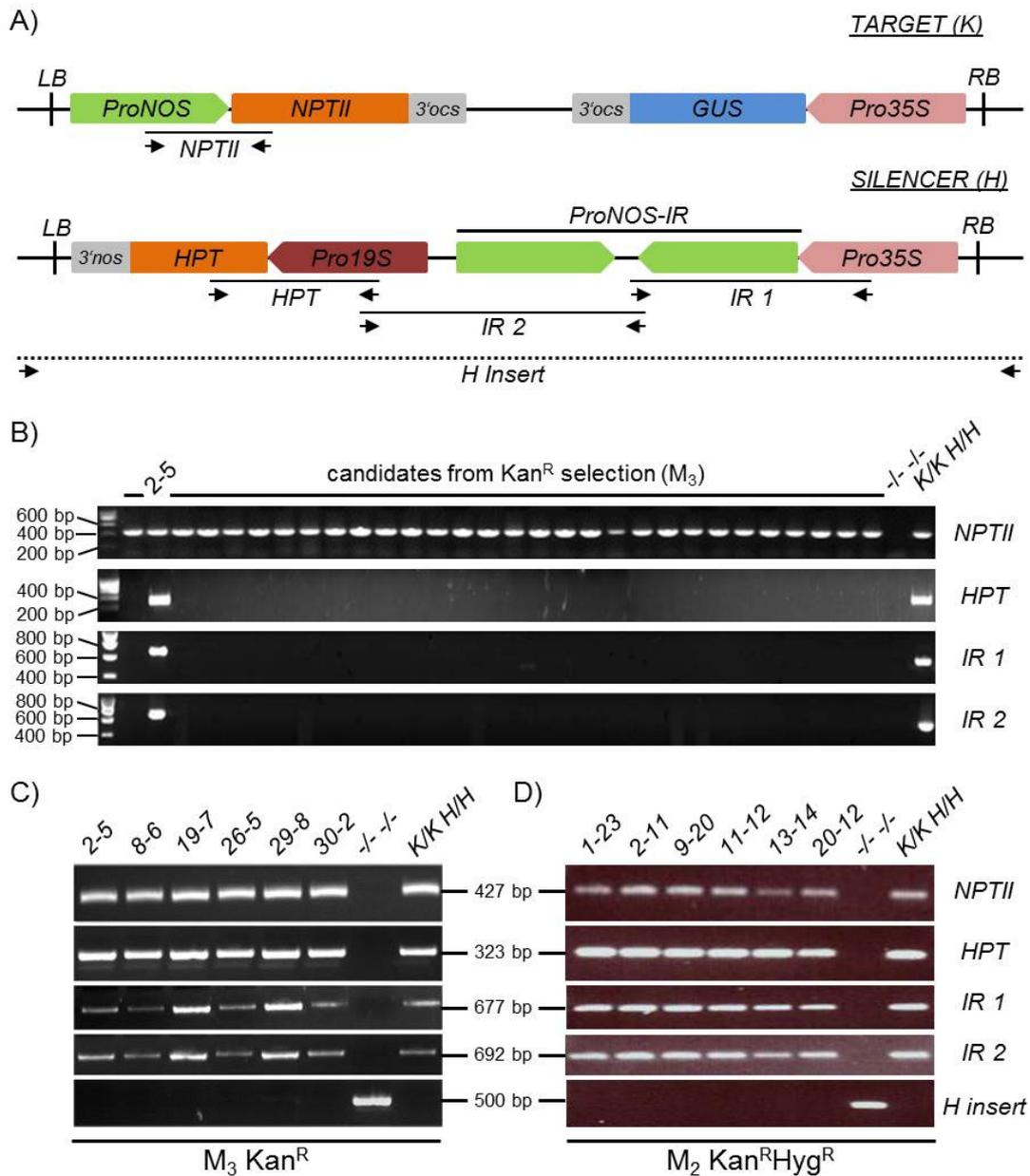
Figure 3: Verification of transgene integrity in candidate mutant lines.

A) Transgene maps. Approximate positions of primers used for transgene-specific PCRs are indicated by arrows. The "H insert" PCR only results in a product when the *H* transgene is absent or heterozygous. B) PCR test for 30 initial mutant candidates derived from the Kan^R screening. C) Transgene integrity in 6 lines showing at minimum 95% Kan^R in M$_3$ generation obtained from Kan^R selection. D) Transgene integrity in mutant lines obtained from Kan^R Hyg^R selection. PCR tests were performed using genomic DNA of original Kan^R Hyg^R M$_2$ plants. PCR product sizes are indicated B), C) Five M$_3$ plants per line were tested to check for segregation of the transgenes, but all gave consistent results.

The result of the first round of screening implied that the seed material submitted to mutagenesis was contaminated with material that has lost the *H* transgene or contained it in a heterozygous manner. Therefore, the majority (98 of 104) of M_2 lines obtained by screening for Kan^R phenotype were "false positive" plants that had lost the *H* transgene by segregation.

As a constitutively expressed *HYGROMYCINE PHOSPHOTRANSFERASE* (*HPT*) reporter gene that confers hygromycin resistance (Hyg^R) is part of the *H* transgene, presence of the *H* can be selected for by germination on medium containing hygromycin in order to minimize the number of "false positive" Kan^R M_2 plants. Thus, M_2 seed stocks were rescreened for individuals that showed a Kan^R Hyg^R phenotype by germination on GM containing 200 mg/l of kanamycin and 20 mg/l of hygromycin. No Hyg^R Kan^R plant was observed in the non-mutagenized C_2 control.

Surviving seedlings grown on agar medium containing hygromycin and kanamycin showed generally weak root growth. To evaluate whether the direct transfer of Kan^R Hyg^R seedlings to soil was possible or a period of recovery at non-selective agar medium was necessary prior to the transfer, two sets of Kan^R Hyg^R M_2 seedlings, 12 per set, from batch no.1 were transferred either directly to soil (*1-11* to *1-22*) or first to non-selective medium (*1-23* to *1-34*). As plant viability was not compromised by the direct transfer to soil, plant recovery at non-selective medium was not further applied. Therefore, the maximum number of transferred Kan^R Hyg^R M_2 plants was reduced to 12 per individual batch (*2-11* to *2-22*; *3-11* and so on). Transferred plants were tested for the presence of both transgenes via PCR. Approximately 110 M_3 seedlings of every obtained Hyg^R Kan^R M_2 plant were tested for Kan^R on GM supplied with 200 mg/l kanamycin. Thirteen M_3 mutant lines (*1-23*, *2-11*, *9-19*, *9-20*, *9-21*, *11-11*, *11-12*, *11-13*, *13-14*, *14-12*, *17-13*, *18-15*, *20-12*) that showed at least 95% Kan^R plants were considered suitable and selected for further analysis.

3.2 THE NPTII PROTEIN LEVEL AS CRITERION FOR "*NO RNA-DIRECTED TRANSCRIPTIONAL GENE SILENCING*" MUTANTS

Amount of NPTII protein in obtained mutant lines

In the transgene system used to perform the genetic screen, compromising of the RdTGS mechanism should result in a release of *NPTII* reporter gene expression. Hence, an increase in NPTII protein as cause for kanamycin resistance should be observed in genuine "*no RNA-directed transcriptional gene silencing*" (*nrd*) mutants. Nevertheless,

studies by others have shown that kanamycin resistance can also arise in *A. thaliana* loss-of-function mutants (Aufsatz *et al.,* 2009; Conte *et al.,* 2009) or gene overexpression lines (Mentewab and Steward, 2005) by affecting chloroplast-localized transporter proteins without any requirement for NPTII expression. Therefore, the amounts of the NPTII protein in relation to the total soluble protein in rosette leaves of mature M_3 plants were determined by ELISA in comparison to control plants grown in parallel (Figure 4).
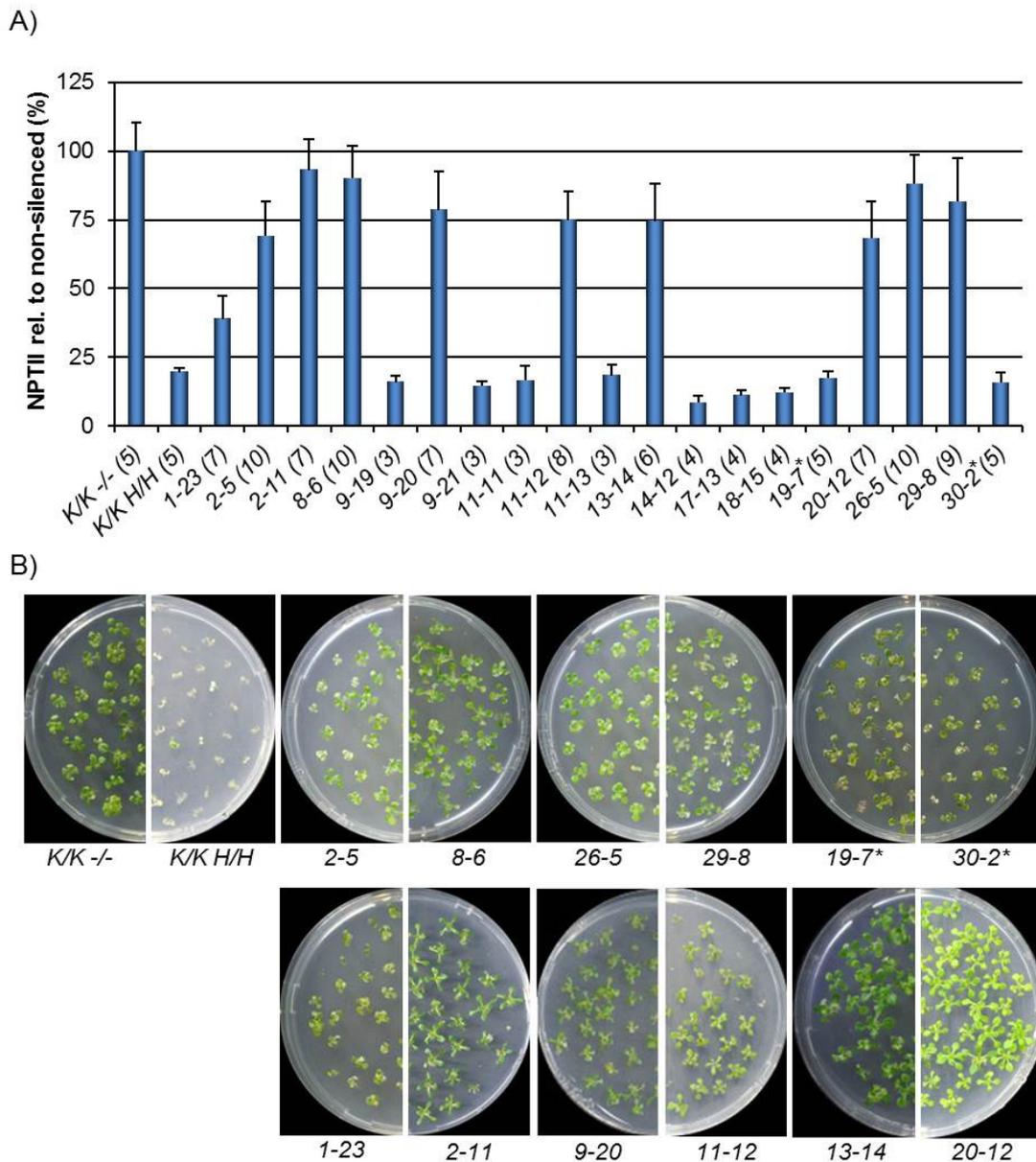
Figure 4: Quantification of NPTII protein levels by ELISA and growth phenotype of M_3 lines

A) Amounts of the NPTII protein were measured in relation to total soluble protein in extracts from leaves of 8-week-old plants. Results are indicated as relative NPTII levels relative to the mean value for non-silenced expression in (*K/K;-/-*) plants (set to 100%). Numbers in parentheses indicate numbers of individual M_3 plants tested. The background signal obtained from non-transgenic control plants was subtracted prior to calculation. Column height represents mean values; error bars represent standard deviation. B) M_3 seedlings of mutant lines showing released NPTII expression compared to mutant lines 19-7 and 30-2 displaying NPTII levels as low as wild-type *K/K;H/H* plants on GM supplied with 200 mg/l kanamycin.

The relative NPTII level was found reduced in non-mutagenized wild type plants harboring both (*K/K*;*H/H*) compared to those that only contained the *K* transgene (*K/K*;-/-), yet not completely abolished. The remaining NPTII might help to explain the high concentration of kanamycin necessary to suppress growth of *K/K*;*H/H* plants in my work in comparison to a previous study utilizing a similar transgene system involving a *NPTII* reporter gene (Aufsatz *et al.,* 2002a; Aufsatz *et al.,* 2002b; Aufsatz *et al.,* 2004).

M$_3$ plants from lines *2-5*, *8-6*, *26-5* and *29-8* from the Kan^R screen and lines *2-11*, *9-20*, *11-12*, *13-14* and *20-12* from the Kan^R Hyg^R screen clearly showed more NPTII protein than *K/K*;*H/H* plants, almost resembling *K/K*;-/- plants. Line *1-23* showed, however less pronounced, also a release of NPTII. Interestingly, despite their Kan^R phenotype, M$_3$ plants from lines *9-19*, *9-21*, *11-11*, *11-13*, *14-12, 17-14, 18-15*, *19-7* and *30-2* did not display a noticeable increase in the NPTII protein compared to wild type *K/K*;*H/H* plants.

3.3   DNA METHYLATION OF THE *TARGET PRONOS* IN *NRD* MUTANTS

In the majority of suppressor of silencing mutants obtained from genetic analysis of transgene-based TGS in *A. thaliana*, the release of silencing is accompanied by a decrease in DNA methylation at the respective transgene. This can be due to impaired siRNA synthesis, inadequate "interpretation" of the siRNA signal or the inability to conduct the methylation at sites addressed by siRNAs (Aufsatz *et al.,* 2002; Kanno *et al.,* 2005; Zheng *et al.,* 2007). Nevertheless, there are mutants like *mom1*, which can cause release of TGS without affecting DNA methylation (Amedeo *et al.,* 2001).

To test for this, DNA methylation at the *ProNOS* in the *K* transgene (*TARGET-ProNOS*) was analyzed in DNA extracted from rosette leaves of individual mature plants (Figure 5B).

Analysis by methylation-sensitive restriction cleavage

The analysis by cytosine methylation-sensitive restriction cleavage coupled to subsequent quantitative PCR allowed testing of the *TARGET-ProNOS* methylation in DNA from multiple individual plants. It revealed a conspicuous amount of amplifiable, and thus uncleaved, genomic DNA in *K/K;H/H* wild type plants after incubation with Psp1406I (cleavage impaired by methylation in CG context) as well as NheI and Alw26I (for both, cleavage impaired in non-CG context) (Figure 5B). The related recognition sites are inside the region covered by the *ProNOS-IR* transcripts and thus these sites are

methylated (Figure 5A). On the other hand, incubation with NcoI targeting a cleavage site outside of the region covered by the *ProNOS-IR* transcripts resulted in almost complete cleavage of the thus unmethylated DNA.

The genomic DNA from M_3 plants of most mutants depicted a preservation of methylation at the Psp1406I site. The strongest reduction in comparison to *K/K;H/H* plants of the same generation was observed for lines *2-11* and *13-14*. Restriction cleavage using NheI and Alw26I indicated low amounts of amplifiable DNA and thus low cytosine methylation at the respective recognitions sites at levels comparable to *K/K;-/-* control plants. Notably, for line *1-23* an intermediate amount of methylation was seen at NheI and Alw26I restriction sites, differing from both, *K/K;H/H* and *K/K;-/-* plants. Consistent with the assumption of persistence of RdTGS of the *NPTII* gene in lines *19-7* and *30-2*, DNA methylation resembling the results of *K/K;H/H* wild type plants was observed for all tested restriction enzymes (Figure 5B).

Bisulfite Sequencing

Analysis using methylation sensitive restriction cleavage could test cytosine methylation only at three sites in the *TARGET-ProNOS*. Determination at single base resolution was performed by bisulfite sequencing. The analyzed part of the top strand of the *ProNOS* in the *ProNOS-NPTII* reporter gene included 69 cytosines, 18 in CG, 12 in CHG and 39 in CHH context, respectively (Figure 5C and Figure S1).

In agreement with the results from restriction cleavage, *K/K;-/-* plants displayed only very few unconverted cytosines and thus at best low cytosine methylation. The observed approx. 1 % of "methylated" cytosines might actually result from incomplete chemical conversion of unmethylated cytosines in the employed protocol than represent genuine methylated cytosines. In contrast to *K/K;-/-* plants, *K/K;H/H* individuals, displayed a high level of DNA methylation evenly distributed along the region of the *ProNOS* sequence covered by the *ProNOS-IR* transcripts at cytosines in all contexts. In particular, 83%, 79% and 59% of cytosine in CG, CHG and CHH context, respectively, were found methylated. Methylation in the CG context was only slightly affected in 6 mutant lines from the Kan[R] screen and somewhat more reduced in lines from the Kan[R] Hyg[R] screen. The CHG and CHH context methylation was clearly reduced in all analyzed mutant lines except *19-7* and *30-2*. It is noteworthy that, in agreement with the restriction cleavage analysis, line 1-23 displayed a less severe decrease of methylation at cytosines in the CHH context. About 25% of cytosines in the CHH context remained methylated in this line (Figure 5C).
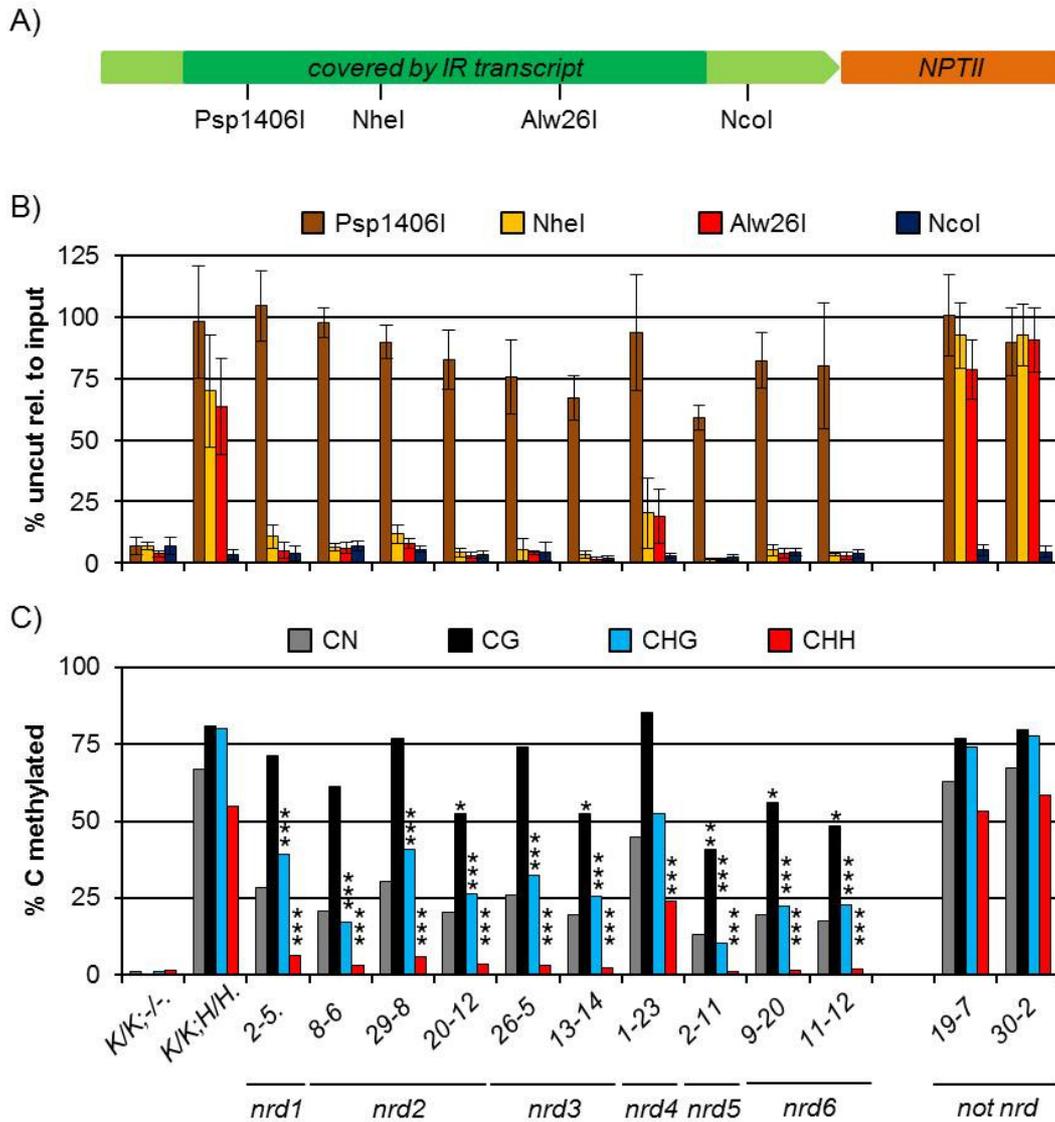
Figure 5: DNA methylation at *TARGET-ProNOS*.

A) Schematic representation and approximal location of the analysed restriction cleavage sites within the *TARGET ProNOS*. B) Cytosine methylation determined by quantitative PCR after cleavage with methylation sensitive restriction enzymes (C in recognition sequence underlined: methylation of cytosine blocks cleavage according to REBASE http://rebase.neb.com) Psp1406I (symetric CG context AACGTT, NheI (CHG and CHH context: GCTAGC), Alw26I (CHH context: GTCTC, GAGAC) and NcoI (CHH context: CCATGG). Results are displayed relative to the mean value for input DNA (set to1). Column hight represents mean values; error bars represent standart deviation. C) Cytosine methylation determined by sodium bisulfite sequencing. Cumulative methylation levels at all cytosines (gray), CG (black), CHG (blue) and CHH (red). H equals C, A or T. For better compareability, mutants were group according their complementation groups determined in later experiments. Exact numbers of analyzed clones and cytosine sites are indicated in Table S7. Asterisks indicate confidence levels (χ^2-test) for being different from *K/K;H/H* values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;

It was expected that kanamycin resistance that is based on the release of TGS of the *ProNOS-NPTII* reporter is accompanied by an increase of *NPTII* transcription and thus increased NPTII protein levels. Only mutant lines matching this criterion were considered as "true positive" *nrd* mutants. However, only 10 out of 19 mutant lines which displayed at least 95% Kan^R plants in M$_3$ progeny possessed elevated NPTII protein levels. The

release of silencing observed in lines *1-23, 2-5, 2-11, 8-6, 9-20, 11-12, 13-14, 20-12, 26-5* and *29-8* was accompanied by reduced DNA methylation in the CHH context at the *TARGET-ProNOS,* indicative for a functioning of the affected genes in the process of RdDM rather than in processes downstream of the establishment of DNA methylation. Furthermore, the reliability of the used two-step strategy, validation of Kan^R resistance in M$_3$ and analysis of NPTII protein levels by ELISA, to exclude "false positive" mutants was proven. Notably, no obvious consistent developmental phenotype was observed in M$_3$ plants of the respective candidates (data not shown).

3.3 DNA METHYLATION OF ENDOGENOUS SEQUENCES

For lines considered to be true RdDM / RdTGS mutants, the DNA methylation status of endogenous target sequences of RdDM was determined by bisulfite sequencing and / or methylation sensitive restriction of genomic DNA and subsequent semi-quantitative PCR in genomic DNA extracted from rosette leaves of mature plants.

In particular, the sequence of the well characterized RdDM target *AtSN1*, a SINE element localized at chromosome 3 and the *INTERGENIC SUBTELOMERIC REPEAT* (*ISR*) of the *MEA* locus (*MEA-ISR*) were analyzed by bisulfite sequencing for all candidates (Myouga *et al.,* 2001; Cao and Jacobsen, 2002; Zilberman *et al.,* 2003). In addition, *AtMU1*, a functional transposable element, was analyzed in most candidate lines in this way (Singer *et al.,* 2001; Lippman *et al.,* 2003). Furthermore, DNA methylation of the less well characterized loci *IGN5*, *IGN23* and *IGN25* (Wierzbicki *et al.,* 2008; Wierzbicki *et al.,* 2009; Bies-Etheve *et al.,* 2009) as well as of the rolling circle transposon derived *BASHO210*, was analyzed in several candidate lines (Hollister and Gout, 2007).

AtSN1

The analyzed sequence of the SINE *AtSN1* is located between the protein coding loci *At3g44000* and *At3g44005* and corresponds to nucleotides 15,794,606 to 15,794,819 of chromosome 3. It contains 4, 7 and 35 cytosines in CG, CHG and CHH context, respectively. Bisulfite sequencing analysis of this sequence showed that approximately 68%, 64% and 29% of cytosines in CG, CHG and CHH context, respectively are methylated in leaf tissue of non-mutagenized *K/K;H/H* plants (Figure 6A). M$_3$ plants of all true *nrd* lines displayed a severe, significant (χ^2-test < 0.01) decrease of overall DNA methylation. As for the *TARGET-ProNOS*, the most prominent decrease was observed for DNA methylation in the CHH context (Figure 6A). This is consistent with mutated loci

being essential for maintenance of DNA methylation in the CHH context, not only of transgenic, but also endogenous RdDM targets. In agreement with the bisulfite data obtained for the *TARGET-ProNOS,* methylation at *AtSN1* in the candidate lines *19-7* and *30-2* was found to be only slightly affected in the CHH context (data not shown). This further supported the assumption that these lines were "false positive" in respect to being affected in RdDM. However, although maintained by mechanistically distinct pathways, also some DNA methylation in CG and CHG context at *AtSN1* was impaired in lines *2-11*, *11-12*, *20-12*, *9-20* and *13-14* as well (Figure 6A).

MEA-ISR

The direct repeats of *MEA-ISR*, a 183 bp sequence located in the intergenic region between the gene loci *At1g02580* and *At1g02590*, were found to be highly methylated in all sequence contexts in vegetative tissue of wild type plants (Cao and Jacobsen, 2002). DNA methylation of the bottom strand of *MEA-ISR*, corresponding to positions 68067 to 68320 of BAC clone T14P4, was assayed (Cao and Jacobsen, 2002). This region contains 9, 2 and 24 cytosines in CG, CHG and CHH context, respectively. Of these, 74%, 23% and 20%, respectively, were found to be methylated in *K/K*;*H/H* plants (Figure 6B).

Similar to *AtSN1*, methylation in CHH context was significantly reduced (χ^2-test, $p<0.05$) or, in case of line *2-11*, completely erased at *MEA-ISR* in all analyzed mutants. Moreover, methylation in the CHG context was also found significantly (χ^2-test, $p<0.05$) reduced or, as in case of lines *2-11*, *8-6*, *13-14*, *26-5* and *29-8*, completely erased in the analyzed mutant lines. In contrast to *AtSN1*, methylation in the CG context was not found significantly reduced (χ^2-test, $p>0,05$) in any of the mutant lines, indicating that the CG context methylation maintenance independent of RdDM is more pronounced at *MEA-ISR* (Figure 6B).

AtMU1

AtMU1, an autonomous Mutator-like DNA transposon present in two copies in the genome of *A. thaliana* accession Col-0, is transcriptionally silenced in *A. thaliana* wild type plants (Le *et al.,* 2000; Singer *et al.,* 2001). In mutants impaired in DNA methylation, such as *ddm1*, *met1*, *hda6* and *cmt3*, a decreased DNA methylation at the terminal inverted repeats (TIRs) of *AtMU1* was observed. This is accompanied by the reactivation of *AtMU1* transcription and transposition, thus linking silencing of this transposon with DNA methylation (Singer *et al.,* 2001; Lippmann *et al.,* 2003). The same study implied

that silencing of *AtMU1* depends on the existence of siRNAs that direct DNA methylation to its TIRs. Further studies clearly depict *AtMU1* as target of the RdDM machinery (Bäurle *et al.,* 2007; Bäurle and Dean, 2008; He *et al.,* 2009).

A possible impairment of the DNA methylation of a 420 bp sequence defined by primers JP1387 and JP1388 at the *AtMU1* copy located at chromosome 4 *(At4g08680)* was analyzed in lines *1-23*; *2-5*; *2-11*; *8-6*; *26-5* and *29-8*. This sequence covers the TIR which is of about 300 bp in length. Per clone 80 cytosine residues were assayed, of which 6 each are in CG and CHG context and 68 are in CHH context.

AtMU1 sequences in wild type *K/K*;*H/H* were found to display 79%, 48% and 24% DNA methylation in CG, CHG and CHH context, respectively. Methylation in the CHH context was found to be significantly reduced (χ^2-test, $p < 0,001$) in all mutant lines analyzed. However, in contrast to the previously analyzed *AtSN1* and *MEA-ISR* sequences, considerable amounts of CHH methylation remained at *AtMU1,* most likely maintained by the partially redundant mechanisms involving *CMT3*. Methylation levels in CG and CHG context were only mildly affected in most lines. Only lines *8-6*, *26-5* and *2-11* displayed significant (χ^2-test, $p < 0,05$) lower DNA methylation in CHG context (Figure 6C).

AtCOPIA4

To address the question whether the obtained mutations also affected DNA methylation that is not dependent on the RdDM mechanism, DNA methylation of the Ty1/copia-like retrotransposon *AtCOPIA4/COPIA-LIKE23 (AtCOPIA4)* was determined by bisulfite sequencing. DNA methylation at *AtCOPIA4* is mainly found in the CG context and is reduced in *ddm1* and *met1* mutants (Johnson *et al.,* 2007; Lippman *et al.,* 2003).

The analyzed sequence region corresponding to the nucleotides 9,485,799 to 9,486,321 of chromosome 4 contained in total 175 cytosines, 25 in CG, 22 in CHG and 128 in CHH context (Figure S2). Approximately 92%, 52% and 9% of cytosines in CG, CHG and CHH context, respectively, were found to be methylated in *K/K*;*H/H* wild type individuals. None of the mutants analyzed displayed significant (χ^2-test, $p < 0.05$) alterations in DNA methylation in any sequence context compared to wild type *K/K*;*H/H* plants (Figure 6D).
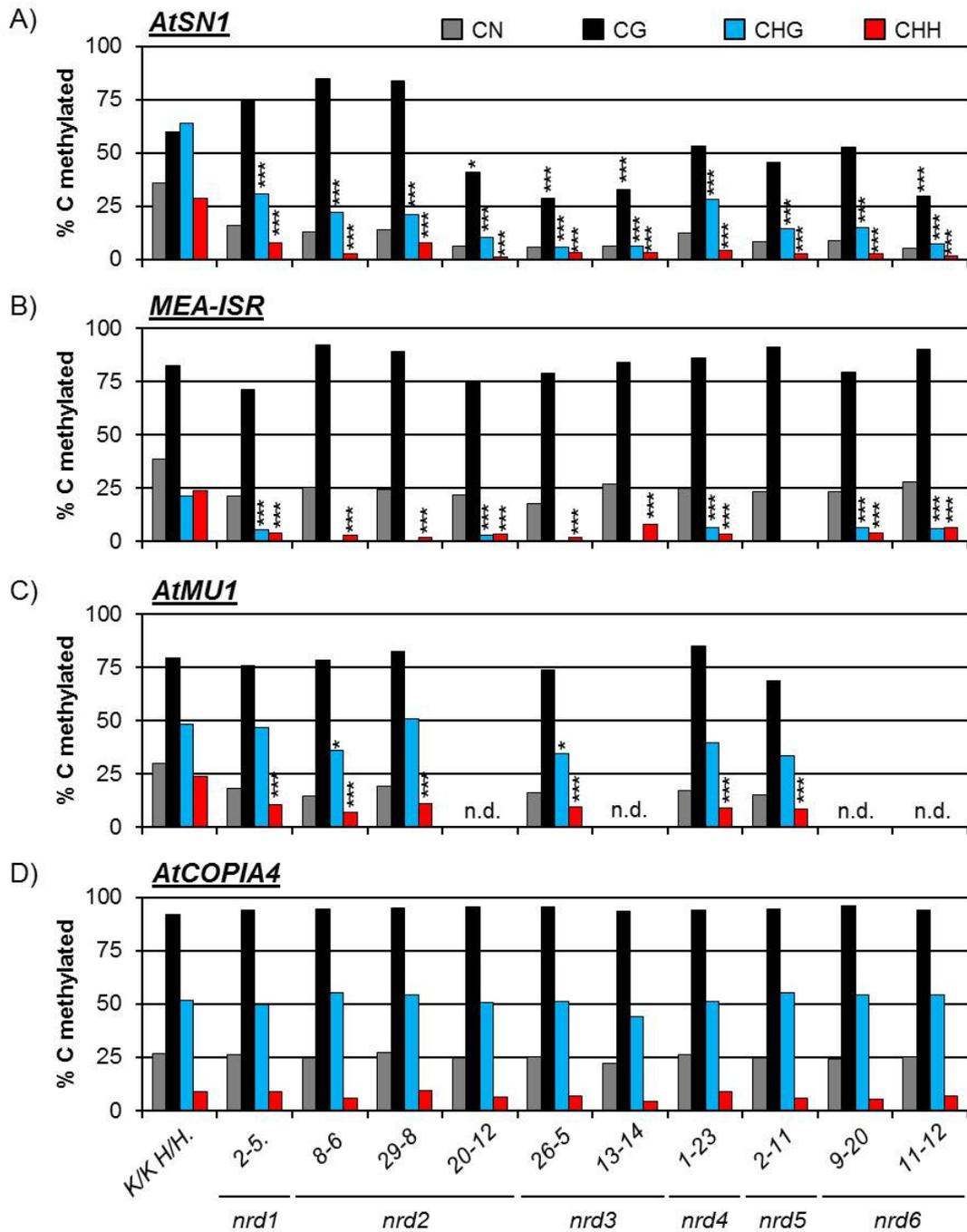
Figure 6: DNA methylation at well characterized endogenous sequences.

DNA methylation patterns of endogenous sequences (A) *AtSN1*, (B) *MEA-ISR*, (C) *AtMU1* and (D) *AtCOPIA4* were analyzed in detail by bisulfite sequencing in non-mutagenized control plants (*K/K*;*H/H*) and isolated mutants. Percentage of cytosine methylation independent of the context (grey columns), in the CG context (black columns), CHG context (blue columns) and CHH context (red columns) are displayed. For better comparability, lines later known to contain allelic mutations are displayed next to each other. Respective putative *nrd* complementation groups are depicted below diagrams. Degrees of significance are indicated. Asterisks indicate confidence levels (χ^2-test) for being different from *K/K*;*H/H* values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Exact numbers of analyzed clones and cytosine sites are indicated in Table S7. n.d. = not determined.

Intergenic non-coding loci

Studies by Wierzbicki *et al.,* (2008) lead to the identification of transcripts from intergenic non-coding regions (*IGN*) in the heterochromatic knob of chromosome 4 (Fransz *et al.,* 2000) that are synthesized by Pol V. AGO4 as well as SPT5L are recruited to chromatin by physical interaction with these transcripts and mediate methylation of the corresponding DNA sequences (Wierzbicki *et al.,* 2009; Bies-Etheve *et al.,* 2009; Rowley *et al.,* 2011). In consequence, DNA methylation at *IGN* sequences is reduced in mutants such as *nrpd2a/nrpe2a, nrpe1*, *drd1*, *dms3* and *rdm1* which are impaired in *IGN* transcription.

DNA methylation level in the CHH context at *IGN5*, *IGN23* and *IGN25* was investigated by methylation-sensitive restriction cleavage using HaeIII followed by semi-quantitative PCR in mutant lines *1-23, 2-5, 2-11, 8-6, 9-20* and *26-5*. Furthermore, bisulfite sequencing of *IGN5* and *IGN23* was performed for selected mutant lines (Figure 7A and Figure 7B). At all three loci, DNA methylation was reduced in M_3 plants of mutants *8-6*, *26-5*, *2-11* and *9-20*. In contrast, methylation of *IGN5* and *IGN23* persisted in *2-5* and methylation at *IGN5-A* and *IGN25* in *1-23*. To confirm previous results, bisulfite sequencing of the analyzed fragments of *IGN5* and *IGN23* was carried out in M_3 individuals of *8-6* and *1-23.*

The analyzed sequence of the top strand of *IGN5* is 210 bp in length and corresponds to nucleotides 2,323,140 to 2,323,350 of chromosome 4 (Wierzbicki *et al.,* 2008). It contains 3, 4 and 33 cytosines in CG, CHG and CHH context, respectively, of which 87%, 79% and 32%, respectively, were found methylated in *K/K*;*H/H* wild type individuals (Figure 7B and Figure S1). In line *1-23*, approximately 85%, 75% and 26% of CG, CHG and CHH context cytosines were found to be methylated, respectively, and no significant reduction in either of the contexts was observed (χ^2-test, p>0.05). In contrast to this result, methylation in all sequence contexts was significantly reduced in line *8-6* (χ^2-test, p<0.001).

The analyzed sequence of *IGN23* is 179 bp in length and corresponds to the coordinates 2,577,896 to 2,578,075 at chromosome 4. It contains 11 CG, 10 CHG and 27 CHH context cytosines, respectively (Figure S1). In wild type individuals, 64%, 38% and 17% of cytosines in the respective contexts were found to be methylated. In contrast, methylation in the CHG context is significantly reduced and in the CHH context almost completely erased in the analyzed mutant lines, whereas the CG methylation is significantly increased in the mutants (χ^2-test, p<0.001) (Figure 7B).

Taken together, the bisulfite sequencing data rather well confirmed the results obtained for these lines by restriction cleavage and PCR.
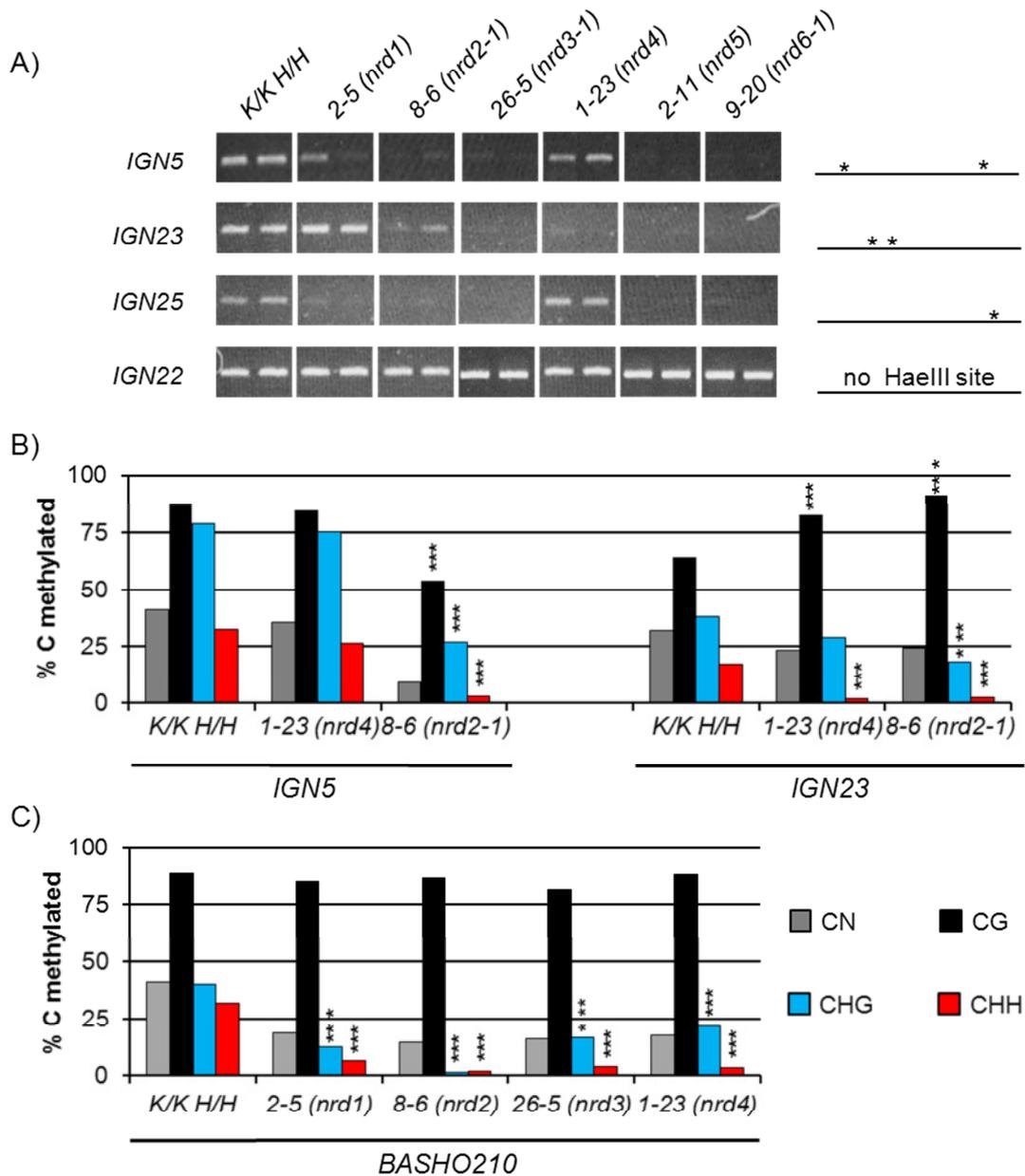
Figure 7: DNA methylation at *IGN* loci and *BASHO210*.

A) Methylation-sensitive restriction cleavage using HaeIII followed by semi-quantitative PCR. DNA preparations from two individual plants per genotype were assayed. The positions of HaeIII restriction sites in *IGN* sequences are marked by asterisks. *IGN22*, which does not contain HaeIII recognition sites, served as loading control. B) Bisulfite sequencing of *IGN5-A* and *IGN23* in M_3 individuals of *8-6* and *1-23*. C) Bisulfite sequencing of *BASHO210* in M_3 individuals of *K/K;H/H*, 2-5, 8-6, 26-5 and 1-23. Degrees of significance are indicated. Asterisks indicate confidence levels (χ^2-test) for being different from *K/K;H/H* values: *** p < 0.001;Numbers of clones and sites analyzed are provided in Table S7.

BASHO210

Helitrons are a recently described class of eukaryotic DNA transposons which are found in high copy numbers in the genomes of *A. thaliana* and *Oryza sativa (O. sativa)* as well as the nematode *Caenorhabditis elegans* (*C. elegans*), and other members of other eukaryote kingdoms (Kapitonov and Jurka, 2007; Cocca *et al.,* 2011). In *A. thaliana*, helitrons and helitron-derived elements represent more than 2% of the genome sequence (Kapitonov and Jurka, 2001). It is assumed that the propagation of helitron elements is conferred by a rolling circle mechanism (Kapitonov and Jurka, 2001). Features of all identified autonomous and non-autonomous helitrons are a TA dinucleotide and a CTRR (R stands for A or G) sequence in their 5' and 3' termini, respectively, the presence of hairpin structures close to their 3' termini and the absence of TIRs. A further hallmark of helitrons is their capability to capture host gene fragments and to transpose these to other locations, potentially causing new protein activities by exon shuffling (Bennetzen *et al.,* 2005; Morgante *et al.,* 2005, Lai *et al.,* 2005).

The *BASHO*-family is a subfamily of non-autonomous helitrons recently described in *A. thaliana* (Hollister and Gaut, 2007). *BASHO210* is a member of subclade V of this family. It is part of the gene coding locus *At2g27070*. In particular, it contributes the 10th exon of the gene. DNA methylation at the 5' end of *BASHO210*, was analyzed in lines *2-5, 8-6, 26-5* and *1-23* as well as in wild type. The analyzed sequence defined by primer *BS-5-B210-F* and *BS-5-B210-R* corresponds to nucleotides 11,555,814 to 11,556,371 of chromosome 2, has a length of 558 bp and contains 9, 4, and 46 cytosines in the CG, CHG and CHH context, respectively (Figure S1). Bisulfite sequencing revealed that 89%, 40% and 32% of cytosines in the CG, CHG and CHH context, respectively are methylated in *K/K*;*H/H* wild type plants. While the CG context methylation was not affected, the CHH methylation was almost completely erased in all analyzed mutants. Furthermore, methylation in the CHG context was found significantly reduced in all analyzed mutant lines. Noteworthy, methylation in this context was found to be almost absent in line *8-6* (Figure 7C).

3.4 *PRONOS* DERIVED SIRNAS

In general, the RdDM pathway can be divided in two major steps: First, the generation of the target specific silencing signal (i.e. the production of the 24 nt siRNAs) and, as the second step, the interpretation of this signal and the establishment of the DNA methylation at the target sequence. To determine if the former or the latter step is

affected in the isolated mutant lines, Northern blot analysis addressing the amounts of 24 nt *ProNOS* siRNAs derived from the *ProNOS-IR* of the *H transgene* was conducted in RNA preparations enriched for "small" RNAs from individual plants of lines *2-5, 2-11, 8-6, 9-20, 11-12, 13-14, 26-5* and *29-8* grown at short day conditions for 8 weeks. In agreement with the absence of DNA methylation at the *TARGET ProNOS*, no *IR*-derived siRNAs were detected in *K/K;-/-* individuals, whereas noticeable amounts were detected in plants containing both transgenes. Moreover, 24 nt *ProNOS-IR*-derived siRNAs were not markedly reduced in any of the analyzed mutant lines if compared to *K/K;H/H* plants, indicating that the mutated factors are not involved in the synthesis or the stabilization of these siRNAs, but rather affects their perception and interpretation (Figure 8).
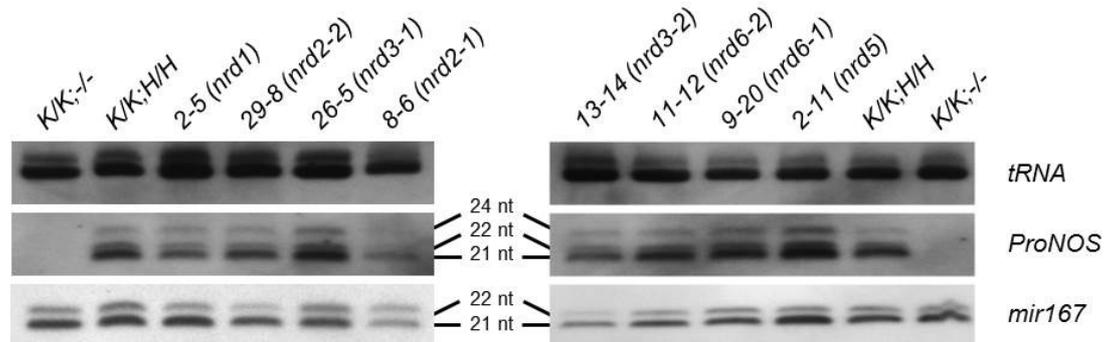
Figure 8: *ProNOS*-derived siRNAs in mutant lines.

Northern blots for siRNA derived from transcripts of the *ProNOS-IR* in the *H* transgene. Blots were hybridized with a RNA probe specific for sense *ProNOS* siRNAs (middle panels). Loading was checked by unspecific hybridization signals of tRNA with the RNA probe (top panels) as well as by re-hybridization with an end-labeled *miR167*-specific probe after stripping (lower panels).

3.5   IDENTIFICATION OF NRD MUTATIONS

Mutations causative for the impairment of RdTGS and accompanying RdDM were identified by a combination of map-based cloning and DNA sequencing techniques, resulting in so far six complementation groups termed *nrd1* to *nrd6*.

3.5.1 IDENTIFICATION METHODS

Map-based cloning

Map-based cloning uses the genetic linkage of a mutation causing a scorable phenotype to sequence polymorphisms at known positions in the genetic or physical map. To determine the position of a mutation, crosses between a mutant in a certain accession and a genetically distinguishable accession are performed. During gametogenesis in the F_1 progeny, stochastically positioned recombination events between homologous chromosomes derived from the parental accessions in meiosis generate "chimeric"

chromosomes. If F_1 plants are then allowed to set seeds by self-pollination, a recessive mutant phenotype will only show up in resulting F_2 individuals that are homozygous for the relevant mutation and, in consequence, contain two copies of the chromosome region surrounding the causative gene locus derived from the one accession that was submitted to mutagenesis. Thus, the position of a mutation can be localized by testing populations of phenotypically mutant F_2 individuals for the positions of recombination events, using e.g. co-dominant markers based on insertion / deletion (InDel-markers) or single nucleotide polymorphisms [Cleaved amplified polymorphisms (CAPS)-markers]. The area containing the relevant mutant gene locus is thus defined by 100% incidence of homozygosity for the marker alleles derived from the respective accession in which the mutation was established; in consequence, the region in which a mutation resides is defined by the closest recombination events known to each flanking side. In this thesis, InDel, CAPS as well as Illumina GoldenGate SNP makers were employed to localize the mutations. Prior to the mapping of mutations, the related genotyping methods where tested in a control F_2 population (C_3F_2) obtained by selfing of F_1 progeny of a cross between *K/K*;*H/H* wild type plants (accession Col-0) and non-transgenic L*er*. As done later by the selection for Kan^R Hyg^R M_3F_2 individuals in the mapping of mutations, the presence of *K* and *H* transgenes in at least the heterozygous state in the C_3F_2 progeny was ensured by first selecting for Hyg^R individuals (presence of *H*) and then for plants positive for ß-glucoronidase activity (GUS$^+$, presence of *K*). The GUS$^+$ criterion was used as the *Pro35S-GUS* reporter gene in *K* is in contrast to the *ProNOS-NPTII* reporter not sensitive to the silencing effect of the *ProNOS-IR*.

Of 220 C_3F_2 individuals screened on GM containing 20 mg/l hygromycin, 152 (69%) displayed a Hyg^R phenotype. Ninety-four of these were further screened for GUS activity, finally resulting in 66 (70%) Hyg^R GUS$^+$ C_3F_2 plants. According to Mendelian segregation, a distribution of 25% homozygosity for the Col-0 allele, 50% heterozygousity and 25% homozygosity for the L*er* allele was expected in this population for all analyzed marker loci that were not genetically linked to *K* or *H* transgenes, while higher incidences of Col-0 alleles were expected for marker loci that were genetically linked to the transgenes.

Genotyping of the Hyg^R GUS$^+$ C_3F_2 population was performed using 20 co-dominant InDel markers. A significant (χ^2-test, $p<0,05$) bias against homozygosity for L*er* alleles for markers *CER461145* and *CER452833* close to the insertion sites of the *K* and *H* transgenes on the lower arms of chromosomes 1 and 4, respectively, was observed (Figure 9). In addition, significant shifts (χ^2-test, $p<0,05$) towards higher incidence of Col-

0 alleles was observed for markers located on chromosome 5 (*CER482932*, *CER450021*, *CER454594*) and lower incidence of Col-0 alleles for markers located on chromosome 2 (*CER460670*, *CER466780*, *CER448739*).

To accelerate genotyping and to increase the marker density, an Illumina GoldenGate assay (GoldenGate assay) allowing multiplex genotyping was developed. As with InDel markers, genotyping of the Hyg^R GUS^+ C_3F_2 population with the GoldenGate assay revealed a significant bias against homozygosity for L*er* alleles for markers physically close to the insertion sites of both transgenes at the lower arms of chromosome 1 and 4, respectively. In agreement with the InDel markers, a significant shift towards low incidences of Col-0 alleles was observed for markers located on chromosome 2 (*ILM2-1, ILM2-2, ILM2-3, ILM2-4*). In addition, a shift towards high incidences of Col-0 alleles was also observed for chromosome 5.
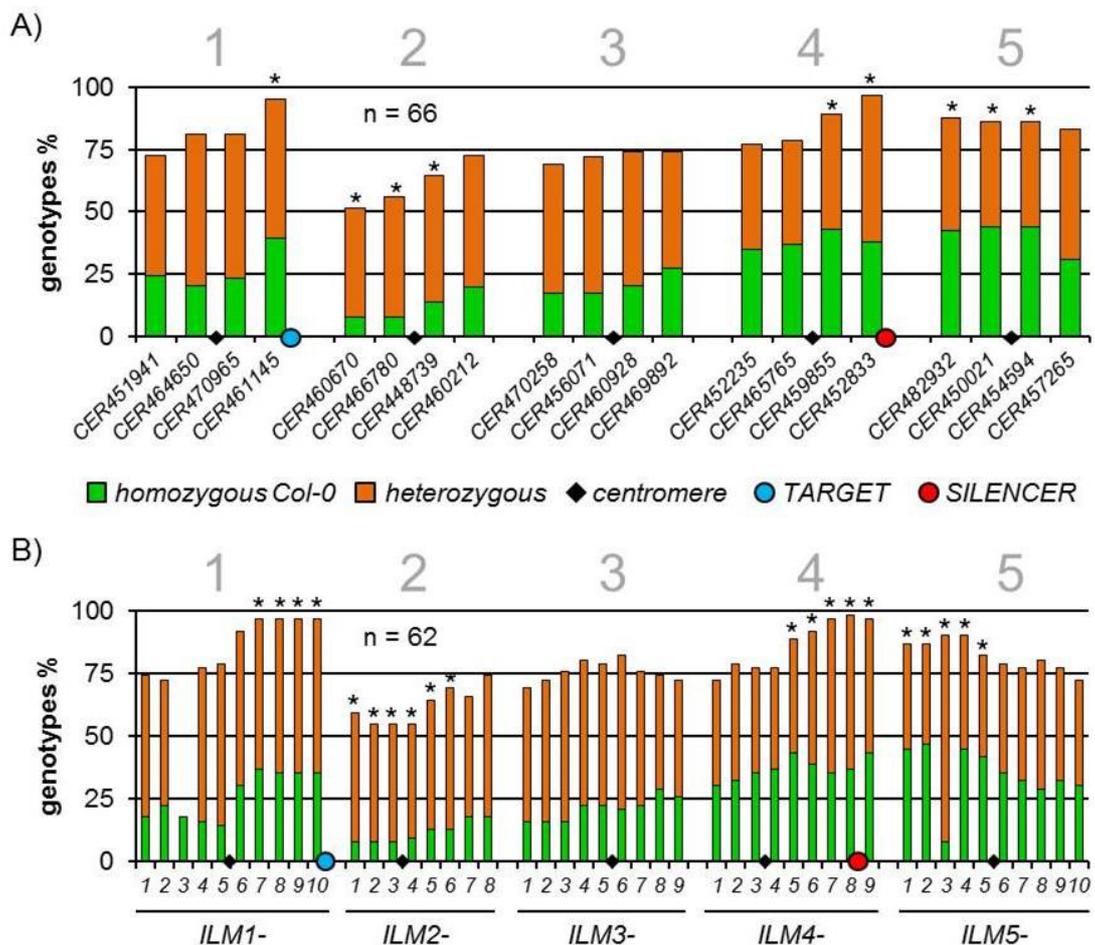
Figure 9: Marker allele incidences in a Hyg^R GUS^+ C_3F_2 population.

Marker allele incidences for A) the initial InDel marker set and B) for Illumina GoldenGate markers. Marker loci showing significant distortion of Mendelian segregation (X^2-test, $p < 0.05$) for any of the three possible genotypes are marked by asterisks. Grey numbers above diagrams refer to chromosomes, marker located on upper / lower arm of respective chromosome are displayed left / right of the centromere indicated by a black diamond. Positions of transgenes are indicated by blue (*K*) and red (*H*) circles.

55

Interestingly, tests with markers *ILM5-2* and *ILM5-4* scored approx. 50% of plants homozygous for the Col-0 allele, while tests for marker *ILM5-3* located between these markers displayed a very low percentage of plants homozygous for the Col-0 allele (Figure 9B). To verify this result, CAPS marker *C5-3443965* (*At5g10920*), *C5-5609978* (*At5g17050*) and *C5-7193938* (*At5g21150*), physically close to *ILM5-2*, *ILM5-3* and *ILM5-4*, respectively, were tested and confirmed the results obtained by the GoldenGate assay (Table S6). Thus, there seem to be a region with apparently distorted recombination in this area. Additional CAPS markers located between *ILM5-2* and *ILM5-4* were tested to estimate the size of the distortion. A comparably high number of recombination events between markers *C5-4165329* (*At5g13120*) and *C5P0888834* (*At5g14020*) and between markers *C5-5609978* (*At5g17050*) and *C5-6086383* (*At5g18370*), respectively, suggested a chromosome area of approximately 1.9 Mb to be affected. This recombination distortion had to be taken into account while determine the positions of the mutations.

Whole genome sequencing

Due to the recent availability of cost effective next generation sequencing technologies, whole genome sequencing (referred to as NGS) of a number of M_3 mutant plants was performed to identify mutations present in their genomic DNA.

As the introduction of the *K* and *H* transgenes into Col-0 plants *via A. tumefaciens* mediated transformation could have had a mutagenic effect by itself, the genome of the used *K/K*;*H/H A. thaliana* wild type line might already have contained sequence deviations from the published Col-0 reference sequence (The Arabidopsis Genome initiative, 2000) prior to submission to EMS mutagenesis. However, due to the stable silencing of the NPTII gene in *K/K*;*H/H* wild type plants, these possible sequence deviations could not be causative for the release of RdTGS and thus had to be excluded in the interpretation of sequence data.

NGS of DNA from a non-mutagenized *K/K*;*H/H* plant yielded 62,130,438 unique reads, of which approximately 96% (59,828,453 reads) were successfully mapped to the *A. thaliana* nuclear or organelle genome sequences. 54,342,788 of these reads corresponded to sequences of the five *A. thaliana* chromosomes, whereas the remaining ones corresponded to mitochondrial and chloroplast DNA. Comparison to the *A. thaliana* reference sequence (TAIR 10) for the nuclear genome identified 2886 variant positions, of which 309 cause non-synonymous mutations in coding sequences (CDS) of annotated *A. thaliana* genes (data not shown).

3.5.2 *NRD1* IS ALLELIC TO *INVOLVED IN DE NOVO 2*

Map based cloning and sequence analysis of *nrd1*

Screening of 305 M_3F_2 plants derived from a cross of mutant M_3 *2-5* with L*er* for $Kan^R Hyg^R$ individuals resulted in the identification of 32 resistant plants. This incidence of approximately 10% resistance plants fits quite well with the 14% expected for a single-locus recessive mutation. Genotyping of the initial mapping population using InDel markers showed a high incidence of homozygosity for the Col-0 allele for marker *CER460928* localized on the lower arm of chromosome 3 (Figure 10A). To exclude occasional "false positive" plants, M_3F_3 progeny obtained by self-pollination of M_3F_2 individuals that were either heterozygous or homozygous for the L*er* allele of the highest scoring marker were tested for the segregation of the $Kan^R Hyg^R$ phenotype. Lines which display markedly less than the expected 56% of $Kan^R Hyg^R M_3F_3$ were removed from the analysis. Using an extended mapping population of 117 $Kan^R Hyg^R M_3F_2$ individuals, the localization of the related *nrd1* mutation on the lower arm of chromosome 3 was further
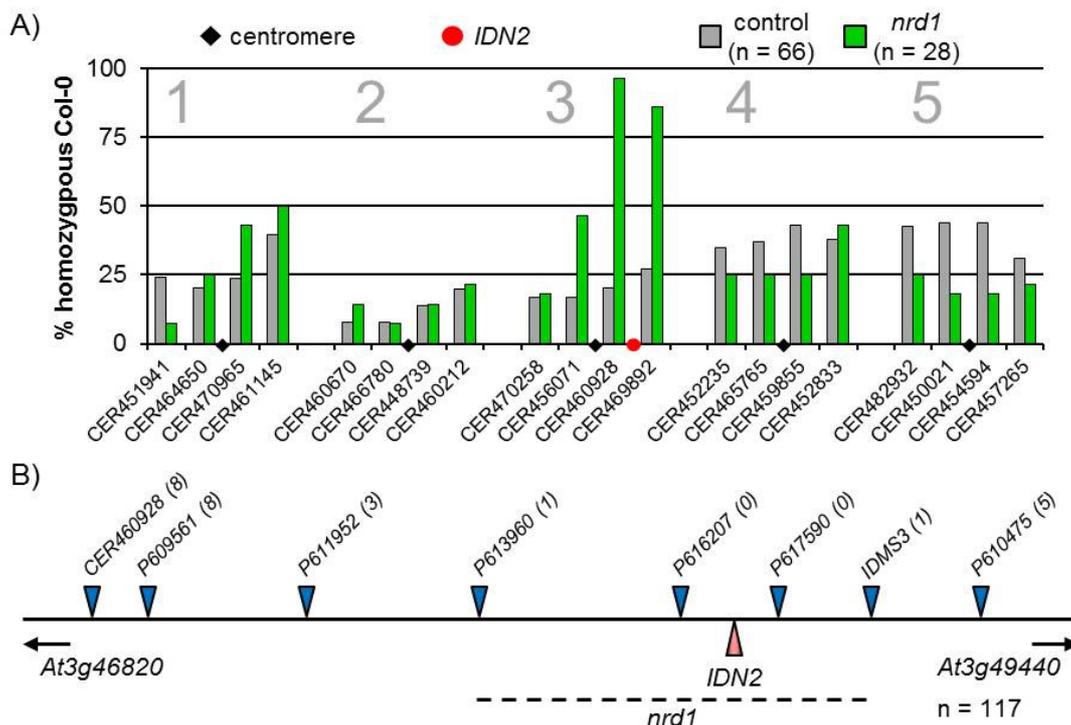
Figure 10: Mapping of *nrd1*

A) Marker allele incidences for the initially tested InDel markers in *nrd1* $Kan^R Hyg^R M_3F_2$ compared to a $Hyg^R GUS^+ C_3F_2$ population. Displayed values are based on an initial mapping population after exclusion of "false positive" plants. Markers located left/right of the centromere are located at the upper/lower arm. The position of *IDN2* is indicated by a red dot. B) Physical map indicating markers and recombination events (numbers in parenthesis) used to delineate position of *nrd1* on the lower arm of chromosome 3. Data are based on a mapping population of 117 plants after removal of "false positive" plants.

mapped to an area defined by recombination events between markers *P613960* and *P616207* as well as *P617590* and *IDMS3*, respectively. These recombination events define an area of approximately 555kb physical distance between *At3g46820* and *At3g49440* (Figure 10B). Sequencing of the gene *At3g48670* which is localized within this region and is coding the previously described RdDM factor *IDN2* revealed a G ➔ A mutation in exon 5 of the gene at position 1883 (Figure 11A). The identified mutation causes the exchange of glycine 514 (G514) for an arginine (R) at protein level (Figure 11B). Protein sequence alignments of IDN2 and its related proteins in *A. thaliana* reveals the evolutionary conservation of G514 (Figure 11C) (Zhang *et al.,* 2012, Xie *et al.,* 2012a, Finke *et al.,* 2012b). Semi-quantitative RT-PCR did not show any notable reduction of *idn2* transcript in *nrd1* compared to *K/K;H/H* wild type plants (Figure 11D). Hence, it can be concluded that G514 is of functional importance and its mutation give rise to a non-functional protein.
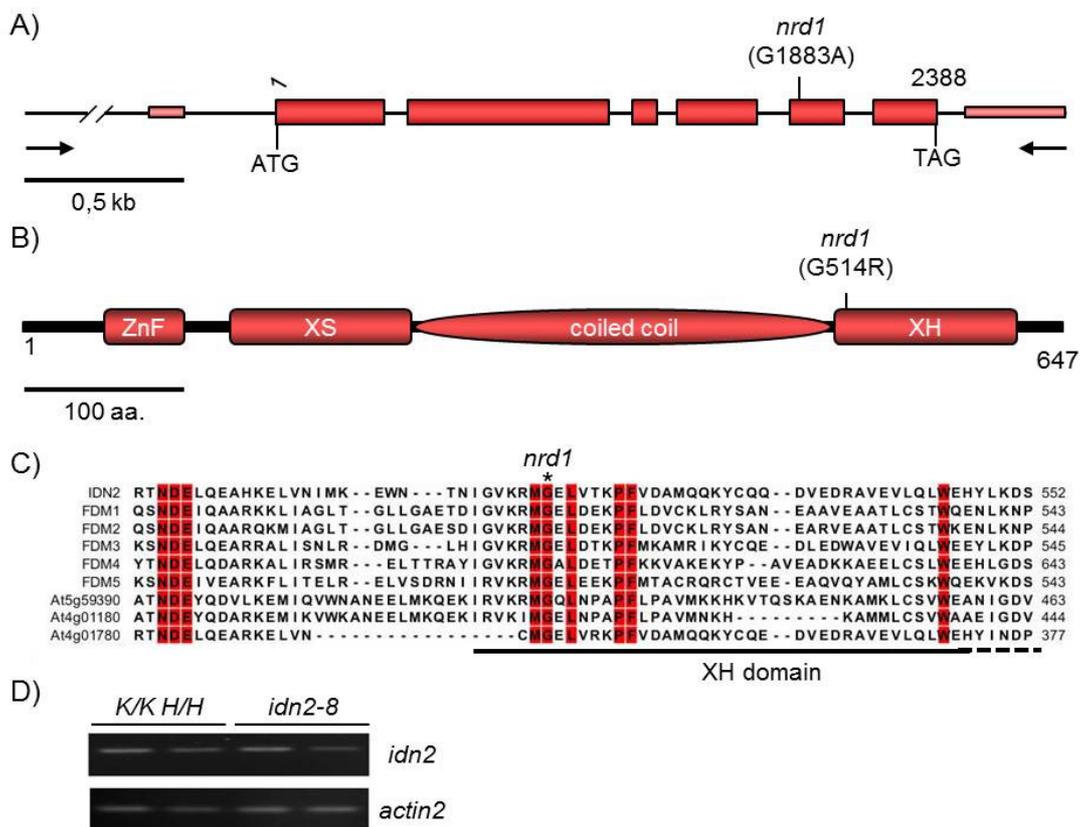
Figure 11: The nature of the mutation in *nrd1*

.A) Gene model of *IDN2 (At3g48670)* indicating the mutation in *nrd1* at position 1883. Numbering is according to the genomic sequence relative to the first nucleotide of the START codon. Arrows indicate the location of the primers used to amplify the wild type gene. B) Protein model of IDN2 indicating the position of the G514R amino acid exchange. The approximate sizes and positions of protein domains are according to the Pfam database (Punta *et al.,* 2012). C) Protein sequence alignment of the XH domain of IDN2 and related proteins in *A. thaliana*. The G residue mutated in *nrd1* is indicated by an asterisk. Sequence alignment was performed using full length protein sequences and ClustalW2 (Larkin *et al.,* 2007) D) Semi-quantitative RT-PCR of *IDN2* and *nrd1/idn2* mRNA. Actin2 was used as reference gene.
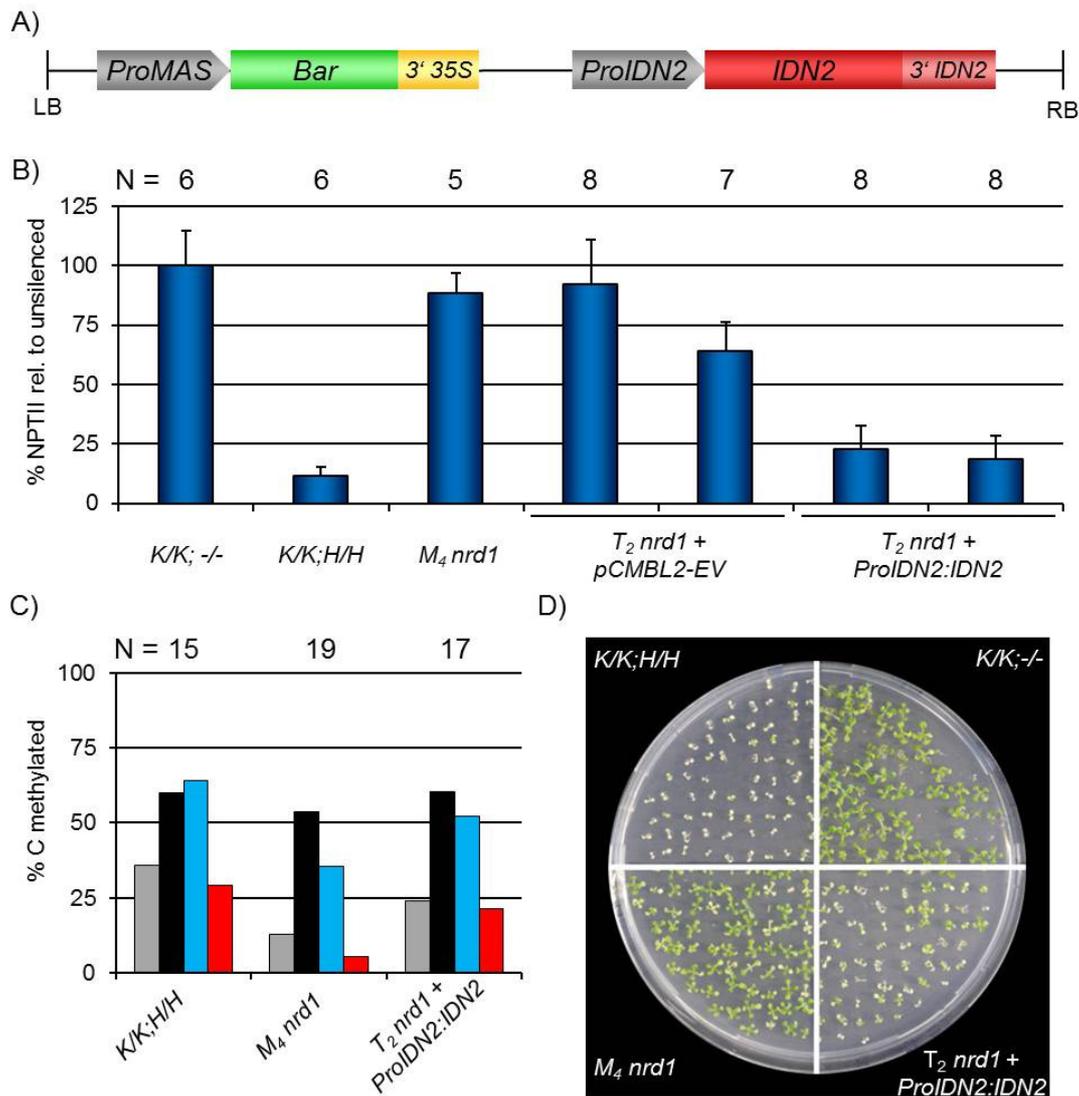
Figure 12: Complementation of *nrd1* by *ProIDN2:IDN2.*

A) T-DNA construct used for complementation of *nrd1* B) NPTII ELISA of T_2 progeny of two independent primary transformants and controls (T_2 progeny of empty vector transformants, *K/K;-/-, K/K;H/H*. M_4 *nrd1*). Column heights represent mean values, error bars represent standard deviation. The number of plants assayed per genotyped is depicted above the diagram C) DNA methylation in CHH context (red columns) at *AtSN1* is reestablished in T_2 individuals after introduction of *ProIDN2:IDN2*. Numbers of clones analyzed are depicted above the diagram C) Segregation of Kan[S] phenotype in T_2 progeny after introduction of *ProIDN2:IDN2*. Approximately 75% of assayed individuals displayed the Kan[S] phenotype indicating the successful complementation of *nrd1*.

Complementation of *nrd1*

To verify that the isolated *idn2* allele is causative for the release of RdTGS in *nrd1*, the coding region of *IDN2* under the control of its native promoter (*ProIDN2:IDN2*) was amplified from genomic DNA of wild type Col-0 via PCR and introduced into *nrd1* M_3 plants by *A. tumefaciens* mediated transformation. In parallel transformation of *nrd1* M_3 plants with the empty binary vector *pCMBL2* (*pCMBL2-EV*) was carried out as a control. The screening of approx. 3×10^5 seedlings germinated from seeds obtained after floral

dip transformation for BASTA resistant (BASTA^R) individuals resulted in 26 primary (T$_1$) transformants.

T$_2$ progeny obtained by self-pollination of five independent T$_1$ individuals was tested for Kan^R on GM containing 200 mg/l kanamycin. In addition, the segregation of the BASTA^R phenotype was assayed in the T$_2$ progeny. In four out of five lines tested, approximately 75% of the T$_2$ progeny showed sensitivity towards kanamycin (Figure 12C). The same held true for the incidence of BASTA^R T$_2$ plants. Furthermore, the *NPTII* expression in T$_2$ progeny of two of the complemented lines was assayed. T$_2$ plants positive for the *ProIDN2:IDN2* displayed severely lower levels of NPTII protein than comparable M$_4$ individuals of *nrd1* and resembled *K/K*;*H/H* individuals (Figure 12A). Moreover, cytosine methylation at *AtSN1* was reestablished in T$_2$ progeny after introduction of *ProIDN2:IDN2* (Figure 12B). In summary, the complementation confirmed that *nrd1* is a new loss-of-function allele of *IDN2.* According to *idn2* alleles described previously (Ausin *et al.,* 2009; Zheng *et al.,* 2010; Xie *et al.,* 2012; Lorkovic *et al.,* 2012), *nrd1* was designated as *idn2-8* (Finke *et al.,* 2012b).

3.5.3 *NRD2* MUTATIONS ARE ALLELIC TO *NRPD2A*

Map based cloning as well as whole genome sequencing approaches revealed that candidates *8-6*, *29-8* and *20-12* are allelic mutants.

Map based cloning

The mutations in candidates *8-6* and *29-8* (*nrd2-1* and *nrd2-2*) were mapped to a region at the upper arm of chromosome 3 using a mapping population of 79 and 67 Kan^R Hyg^R M$_3$F$_2$ plants, respectively (Figure 13). The region spanned approx. 829kb physical distance and was defined by recombination events between markers *C3AB015474* and *C3P0484614* as well as between marker *MN38693286* and *CER456071*. Sequencing of *NRPD2a (At3g23780)*, a gene locus within this region known to be involved in RdDM, revealed G ➜ A mutations in exon 2 at position 1590 (*nrd2-2*) and exon 7 at position 5977 (*nrd2-1*) of *NRPD2a*. This locus encodes the common second-largest subunit of Pol IV and V (Herr *et al.,* 2005; Kanno *et al.,* 2005; Onodera *et al.,* 2005; Pontier *et al.,* 2005; Ream *et al.,* 2009). While the mutation in *nrd2-2* causes the exchange of a glycine for an aspartate at position 174 (G174D) of the protein, the mutation identified in *nrd2-1* leads to the exchange of a glutamate for a lysine at position 1079 (E1079K).
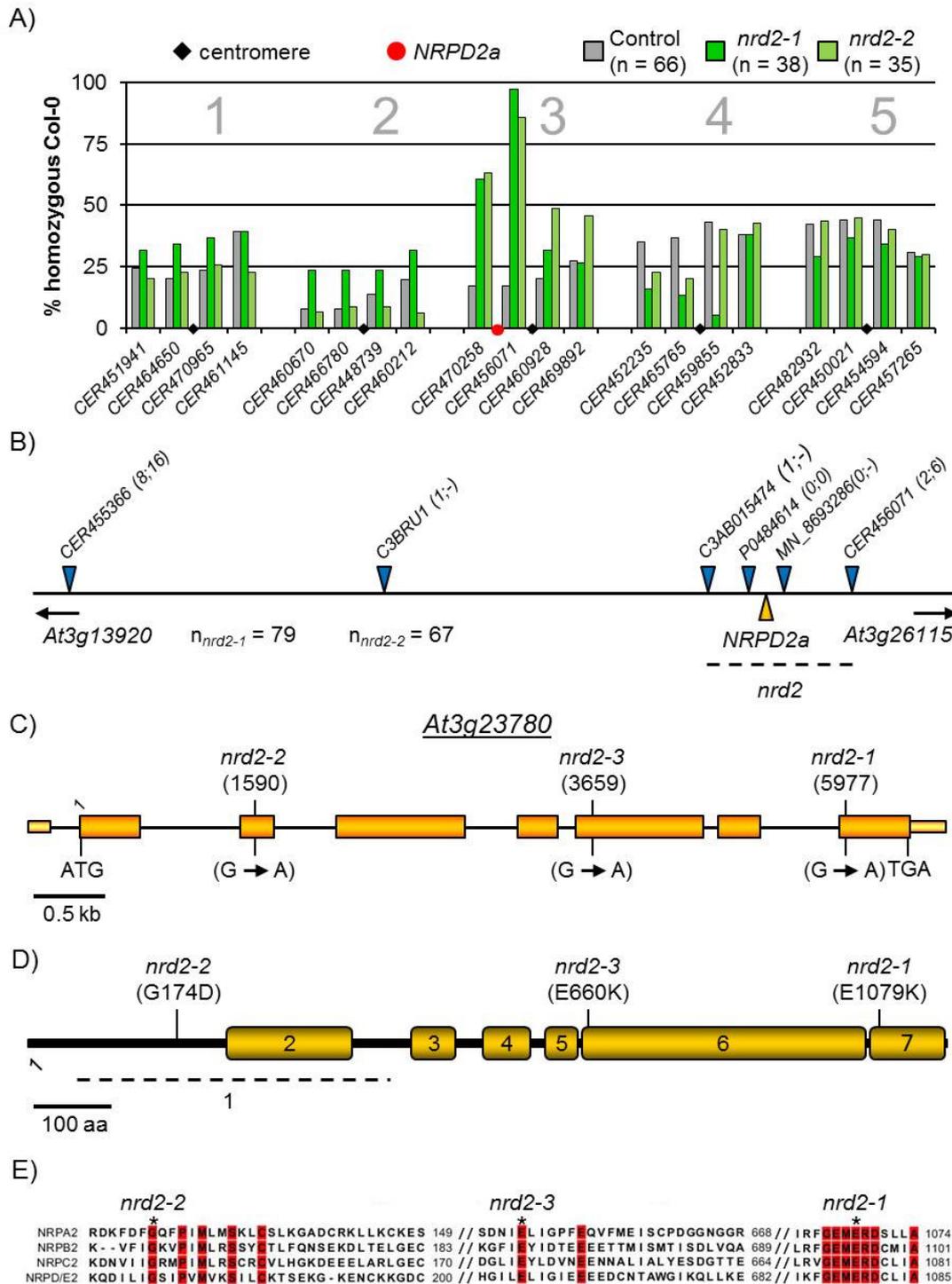
Figure 13: *nrd2*

A) Marker allele incidences among the initial set of InDel markers in Kan^R Hyg^R M$_3$F$_2$ populations derived from *nrd2-1* and *nrd2-2* after of "false positive" individuals. B) Fine mapping. Recombination events are depicted in parentheses for *nrd2-1* and *nrd2-2*, respectively. C) Gene model of *NRPD2a*. Positions and base pair exchanges for identified *nrd2* alleles are depicted above sequence. Numbering is relative to the first nucleotide of the START codon. D) Protein model of NRPD2a. Positions and nature of amino acid exchanges are displayed. Size and position of conserved protein domains are displayed according to the Pfam database (Punta *et al.,* 2012). E) Protein sequence alignment of second largest subunits of *A. thaliana* NRPs. Evolutionary conserved residues are highlighted in red. Residues affected in *nrd2* mutants are marked by asterisks. For alignment, full length protein sequences and ClustalW2 (Larkin *et al.,* 2007) were used.

Sequence alignments of NRPD2a-related second-largest subunits of DNA dependent RNA polymerases (Pol I, Pol II, Pol III) of *A. thaliana* and of other eu- and prokaryotic organisms revealed an evolutionary conservation of the affected residues in all of these proteins (Figure 13 and Figure S3). Hence, it was concluded that these amino acid residues were of functional importance and that the respective altered proteins are non-functional.

Complementation test using *nrd2-1 and nrd2-2*

To proof that the *nrpd2a* alleles identified in *nrd2-1* and *nrd2-2* cause the observed release of RdTGS and RdDM, crosses between these alleles were carried out and the amount of the NPTII protein was determined by ELISA in F_1 individuals grown under a long day regime for 5 weeks. The F_1 progeny of crosses between *nrd2-1* and *nrd2-2* display clearly more NPTII protein compared to the F_1 progeny of control crosses performed between *K/K*;*H/H* wild type plants and *nrd2-1* or *nrd2-2* (Figure 14A). Allelism was further supported by persistent low DNA methylation at the *AtMU1* locus in F_1 plants from the *nrd2-1* x *nrd2-2* crosses, which resembled the DNA methylation in the M_4 progeny of *nrd2-1.* Interestingly, *AtMU1* methylation in the F_1 progeny of the control crosses of *K/K*;*H/H* wild type plants and *nrd2-1* or *nrd2-2* seemed to be reestablished, but did not immediately revert to initial wild type levels (Figure 14B).

In addition to the analysis of the F_1 generation, segregation of the Kan^R phenotype in F_2 progeny was determined. Almost all (237 of 250; 95%) F_2 seedlings derived from non-complementing F_1 displayed Kan^R phenotypes. In contrast, only 19% (47 of 253) and 15% (35 of 241), respectively, were found to be Kan^R in F_2 progeny of the control crosses of *nrd2-1* or *nrd2-2* to *K/K*;*H/H* plants containing functional NRPD2a, coming rather close to the expected 25% (Figure 14C).

Thus, *nrd2-1* and *nrd2-2* were confirmed to be *nrpd2a* loss-of-function mutants. According to the recent numbering they were designated as alleles *nrpd2a-54* and *nrpd2a-55* (Lopez *et al.,* 2011; Finke *et al.,* 2012b).

Identification of *nrd2-3* by whole genome sequencing

In *20-12* (*nrd2-3*) a mutation in the *NRPD2a* gene was detected by next generation sequencing (NGS). In total, 45,562,872 reads were obtained of which 40,715,378 reads were properly mapped to sequences in the five chromosomes of *A. thaliana*. On average every base pair was covered by 32 reads. Data analysis resulted in the identification of 1086 SNPs in the sequenced M_3 plant of *nrd2-3* that were not present in the sequenced
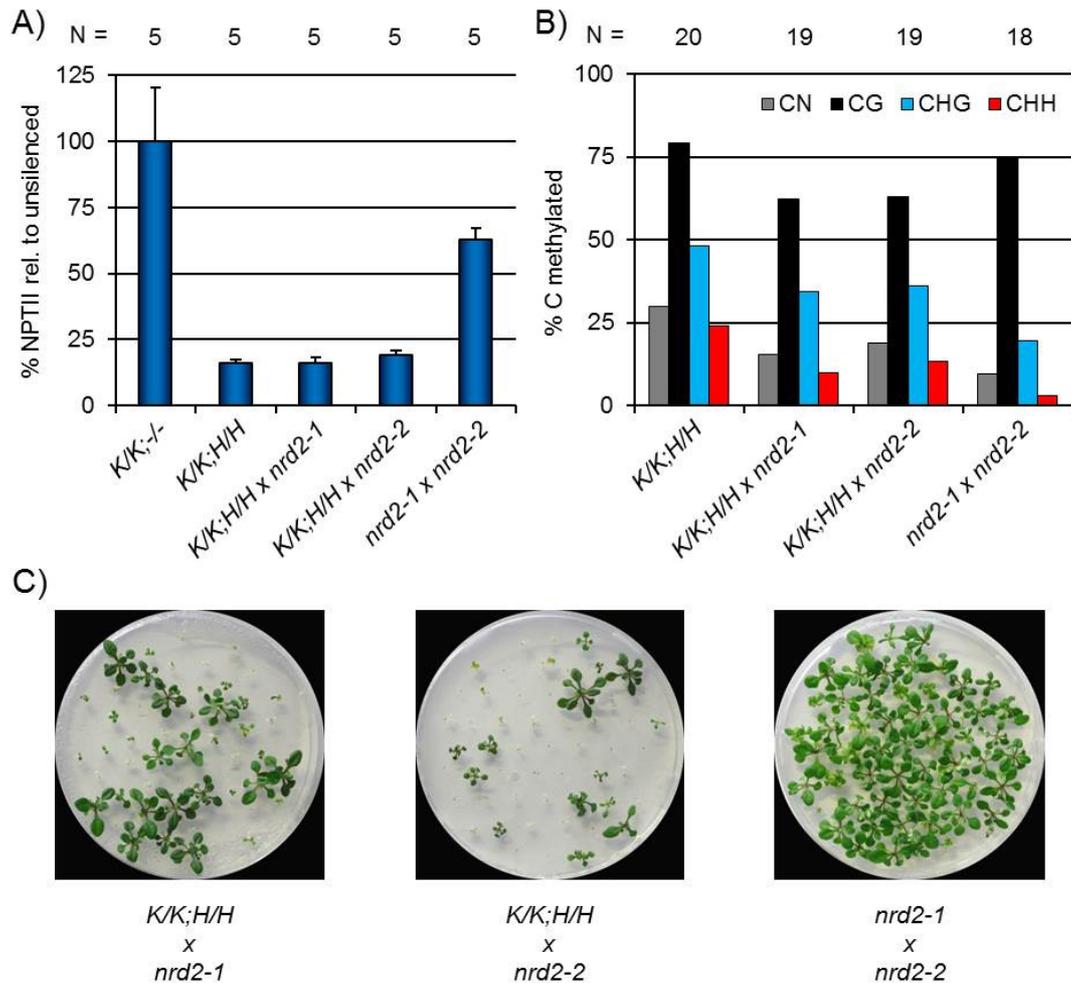
Figure 14: Lack of complementation between *nrd2-1* and *nrd2-2*

A) NPTII ELISA of F$_1$ progeny of a complementing crosses between *nrd2-1* and *nrd2-2*. Values for comparable *K/K;-/-*, *K/K;H/H* and of F$_1$ progeny heterozygous for *nrd2* alleles after back crosses to wild type plants are displayed as controls. Protein levels were determined in leaves of 5-week-old plants Column heights represent mean value, error bars represent standard deviation. The number of plants analyzed is depicted above the diagram. B) DNA methylation at *AtMU1* in F$_1$ progeny of backcrosses and crosses between *nrd2-1* and *nrd2-2*. The number of clones analyzed per genotype is indicated. C) Segregation of Kan[R] phenotype in F$_2$ individuals of after backcrosses (left and mid) and cross of *nrd2-1* and *nrd2-2*.

K/K;*H/H* individual. Of these SNPs, 132 were found to cause homozygous, non-synonymous mutations in the CDS of protein coding genes, including a G ➔ A mutation at position 3659 in exon 4 of *NRPD2a*. This mutation causes the exchange of glutamate 660 for a lysine (Figure 13E) at protein level. Similar to the previously isolated *nrpd2a* alleles, the affected residue is evolutionary conserve among the second-largest subunits of DNA dependent RNA polymerases in *A. thaliana* as well as in those of other eukaryotic organisms (Figure 13 and Figure S3), suggesting a functional importance of the residue. Therefore, candidate *20-12* is considered to harbor another loss-of-function allele of *nrpd2a* (*nrpd2a-af*) and thus belongs to the *nrd2* complementation group.

However, the genetic verification of this conclusion was not possible due to time constrains within the frame of this thesis.

3.5.4 *NRD3* MUTATIONS ARE ALLELIC TO *DRM2*

Map based cloning approach

To identify the mutation causative for the release of RdTGS in mutant *26-5*, an initial population of 317 M_3F_2 plants was tested on selective GM resulting in an initial mapping population of 34 Hyg^R Kan^R M_3F_2, individuals. After removal of "false positive" M_3F_2 by analysis of the incidence of the Hyg^R Kan^R phenotype in M_3F_3 progeny 32 "true positive" plants remained. These 32 true Hyg^R Kan^R M_3F_2 plants were genotyped using the GoldenGate assay. Genotyping resulted in the identification of several marker loci located at different chromosomes that were homozygous for the respective Col-0 alleles in a high proportion of the population (Figure 15A). In particular, all plants were found to be homozygous for Col-0 alleles at markers *ILM2-2, ILM2-3* and *ILM2-4,*. Furthermore, also at markers *ILM5-2, ILM5-3* and *ILM5-4*, all plants were homozygous for Col-0 alleles (Figure 15A). A mapping population of in total 76 true Hyg^R Kan^R M_3F_2 plants was genotyped with additional markers for chromosome 2 and 5, to further delimitate the areas homozygous for Col-0. This resulted in the identification of a section at chromosome 2 defined by markers *2g02770* (*At2g02770*) and *C2P312504* (*At2g11970*) that covered an area of approx. 4.2 Mb of physical distance (19 cM genetic distance) (data not shown) and a section of 5.1 Mb (23 cM) defined by markers *CER482932* (*At5g06750*) and *C5-7193938* (*At5g21150*) at chromosome 5 (Figure 15B). As mutant *26-5* was mapped to regions different from *nrd1* and *nrd2*, it was assigned to a new complementation group as *nrd3-1*. As the 76 true Hyg^R Kan^R M_3F_2 plants, represent approximately 10% (76/779) of total screened individuals the presence of a single-locus recessive mutation was anticipated.

No gene involved in RdDM and RdTGS was described so far for the area defined by *2g02770* and *C2P312504* on chromosome 2. However, two gene loci, *DRM2* (*At5g14620*) and *DRM1* (*At5g15380*), relevant for RdDM, are localized between *CER482932* and *C5-7193938* on chromosome 5. DRM2 is the DNA methyltransferase predominantly responsible for *de novo* DNA methylation and maintenance of methylation in the CHH context in *A. thaliana* (Cao and Jacobsen, 2003), while *DRM1* encodes a DRM2-homolog specifically expressed in the egg cell. Although both proteins seem to harbor all elements required for DNA methyltransferase activity, only mutations in *DRM2*

results in a significant reduction of DNA methylation in mature *A. thaliana* leafs (Cao and Jacobsen, 2003; Julien *et al.,* 2012).
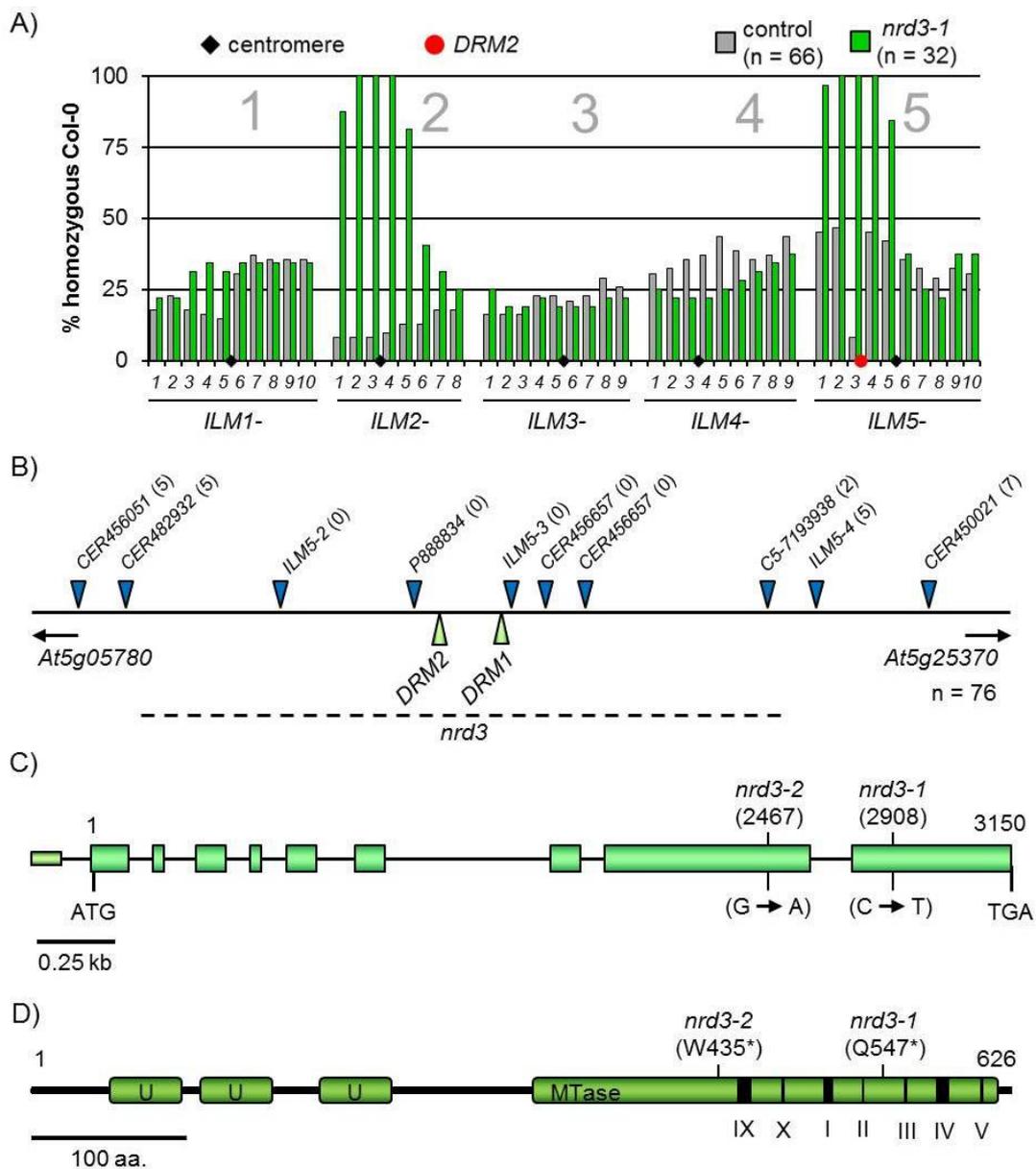
Figure 15: *nrd3*

A) Marker allele incidences in a Kan[R] Hyg[R] M_3F_2 mapping population derived from *nrd3-1* as determined by the Illumina GoldenGate assay. B) Fine mapping of the position of *nrd3-1*. Numbers of recombination events among 76 Kan[R] Hyg[R] M_3F_2 detected by Indel and CAPS markers are indicated in parenthesis. C) Gene model of *DRM2* (At5g14620). Position of identified nucleotide exchanges in mutant alleles for *nrd3-1* and *nrd3-2* are displayed in parenthesis. Numbers refer to the position in the genomic sequence relative to the first nucleotide of the START codon. D) Protein model of DRM2. Positions of premature STOP codons are indicated above symbolized by asterisks. Annotation of UBA domains (U), C-5 cytosine methyltransferase domain (MTase) are displayed. Position of conserved catalytic residues are highlighted as black bars and numbered according to established nomenclature (Henderson *et al.,* 2010).

Sequencing of PCR products amplified from *At5g14620* in five M_3 individuals of *nrd3-1* lead to the identification of a C ➔ T transition in exon 9 at position 2908 causing a nonsense mutation and thus a STOP codon at position 547 of the protein (Figure 15C and D). The resulting truncated protein lacks motives III, IV and V of the methyltansferase domain, which are highly conserved among DNA methyltransferases (Chang et al., 1995; Cao et al., 2000). As mutation of *DRM2* is sufficient to impair *de novo* methylation of transgenes and endogenous loci and the incidence of the $Kan^R Hyg^R$ phenotype is consistent with a single causative mutant locus, it is plausible that no further mutation affecting RdTGS and RdDM is present in candidate *26-5* and that the linkage to chromosome 2 observed in marker-based mapping is an artifact, possibly due to chromosomal rearrangements.

Complementation of *nrd3-1* with transgenic *DRM2*

To verify that the causative mutation in *nrd3-1* is allelic to *drm2*, the coding region of *DRM2* under control of its native promoter (*ProDRM2:DRM2*) was amplified from genomic DNA of wild type *A. thaliana* Col-0 *via* PCR and introduced into *nrd3-1* M_3 plants via *A. tumefaciens* mediated transformation (Figure 16A). In parallel, transformation of *nrd3-1* M_3 plants with the empty binary vector *pCMBL2* (*pCMBL2-EV*) was carried out as control.

Approximately 3×10^4 (*ProDRM2:DRM2*) and 1.5×10^4 (*pCMBL2-EV*), respectively, seeds obtained from floral dip transformation were screened for $BASTA^R$ T_1 transformants. The isolated $BASTA^R$ plants were tested for the presence of the three transgenes via PCR using specific primers. The segregation of the Kan^S phenotype in T_2 progeny of independent primary transformants was assayed for three of the *ProDRM2:DRM2* lines and two of the *pCMBL2-EV* lines obtained. Approximately 220 T_2 seeds per line were germinated on GM containing 200 mg/l kanamycin. For two of the *ProDRM2:DRM2* lines, approx. 75 % of the T_2 seedlings displayed a Kan^S phenotype matching the expected value for a successful complementation by a single locus insertion, whereas the T_2 progeny of the empty vector lines displayed approx. 100% Kan^R phenotype (Figure 16D). Furthermore the amount of NPTII protein in leaves of mature T_2 plants was analyzed in comparison to *nrd3-1* M_4 mutant plants. T_2 progeny grown on soil was screened by PCR for the presence of the *BAR* resistance marker. The amount of NPTII protein in *ProDRM2:DRM2* plants resembled the values found in *K/K;H/H* wild type plants, whereas plants transformed with the empty T-DNA vector displayed NPTII levels similar to M_4 progeny of *nrd3-1* (Figure 16C).
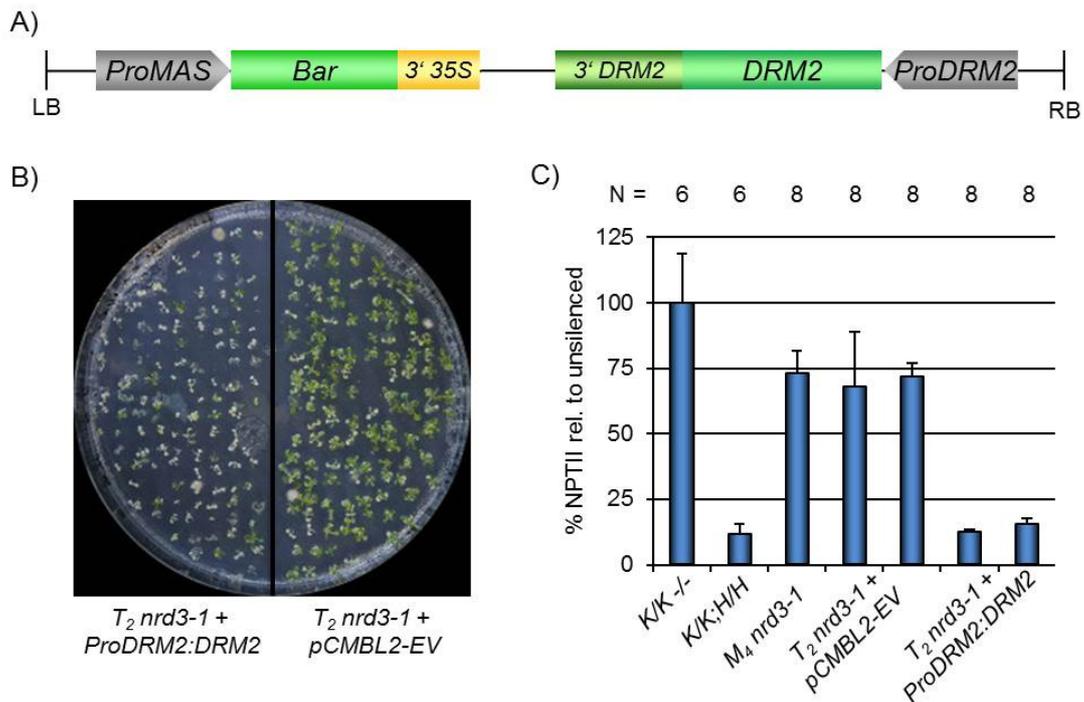
Figure 16: Complementation of *nrd3-1* by *ProDRM2:DRM2*.

A) T-DNA construct used for complementation of *nrd3-1* B) Segregation of Kan[S] phenotype in T$_2$ progeny after introduction of *ProDRM2:DRM2*. Approximately 75% of assayed individuals display the Kan[S] phenotype indicating reestablishment of *ProNOS-NPTII* RdTGS and thus the successful complementation of *nrd3-1*. C) NPTII ELISA of T$_2$ progeny of two independent primary transformants with *ProDRM2:DRM2*, controls T$_2$ progeny of empty vector transformants, *K/K;-/-, K/K;H/H* and M$_4$ *nrd3-1*. Column heights represent mean values, error bars represent standard deviations. Protein levels were determined in leaves of 5-week-old plants. Number of plants assayed is indicated above diagram.

Identification of *nrd3-2* by whole genome sequencing

In *13-14*, a mutation in *DRM2* was detected by NGS of DNA from one M$_3$ plant. In total 47,194,081 reads were obtained, of which 30,158,845 (63.9%) were properly mapped to the *A. thaliana* nuclear genome. On average, every base pair was covered by 22 reads. Analysis of these reads, led to the identification of 3217 SNPs differing from the *A. thaliana* reference sequence. Among the 1071 SNPs exclusively found in mutant *13-14*, 80 were found to cause non-synonymous, homozygous mutations in the CDS of protein coding genes. One of these mutations was found at position 2467 (relative to the 1st nucleotide of the START codon) in exon 9 of *DRM2*. The putative nonsense mutation results in a premature STOP codon at position 435 of the protein. Therefore, the predicted protein is lacking all domains conserved in DNA methyltransferases. The presence and homozygous nature of the mutation in M$_3$ progeny was confirmed by Sanger sequencing of PCR products obtained from *DRM2* in five M$_3$ individuals of line *13-14*. Hence, the causative mutation in *13-14* is very likely allelic to *drm2* and was assigned as second member *nrd3-2* to complementation group *nrd3*. As no verification

by complementing crosses between *nrd3-1* and *nrd3-2* or by transformation with *ProDRM2:DRM2* could be carried out in the context of this thesis due to lack of time, the putative *drm2* loss-of-function allele was designated as *drm2-af*.

3.5.5 *NRD4* IS POTENTIALLY ALLELIC TO AGO6

Map based cloning approach

Screening of 634 M_3F_2 plants derived from mutant 1-23 for Kan^R Hyg^R individuals resulted in an initial mapping population of 42 plants. The Genotyping of this population with the help of the GoldenGate assay detected a very high proportion of plants homozygous for Col-0 alleles at markers *ILM5-2* and *ILM5-4* on the upper arm of chromosome 5 (Figure 17). Further genotyping of a total of 128 Kan^R Hyg^R M_3F_2 individuals with chromosome 5-specific CAPS marker supported these results. As in case of the Hyg^R GUS^+ C_3F_2 control population, significantly fewer plants homozygous for the Col-0 allele were found for marker *ILM5-3* than for the flanking markers *ILM5-2* and *ILM5-4*. However, the difference was more pronounced than in the control population. Segregation of the Kan^R Hyg^R phenotype was assayed in M_3F_3 progeny of individuals heterozygous for all markers tested at chromosome 5. All such individuals were found to be "false positive" and withdrawn from further analysis, resulting in a final mapping population of 117 plants, representing approximately 7.3% of the 1603 plants initially screened for the Kan^R Hyg^R phenotype. This differs significantly (χ^2-test, p>0,01) from the incidence of 14% Kan^R Hyg^R individuals expected if a single recessive mutation is causative for the resistance phenotype. Sequencing of *DRM2* and *DRM1*, the only loci known to be involved in RdDM located between *ILM5-2* and *ILM5-4*, did not lead to the discovery of any mutations in the CDS or the promoter of these genes. Therefore, it was assumed that a locus previously not described to be involved in RdDM is causative for the release of RdTGS in mutant *1-23*. Thus, 1-23 was assigned to the new complementation group *nrd4*. However, a lower peak of enhanced incidence for Col-0 alleles in Kan^R Hyg^R M_3F_2 was also observed around marker *ILM2-7* on the lower arm of chromosome 2.

Whole genome sequencing

In parallel, it was attempted to identify the mutated locus causative for the release of the NPTII silencing in *nrd4* by NGS. The Sequencing of the DNA from one M_3 individual of *nrd4* resulted in 69,416,025 reads of which 63,765,878 (91.9%) reads were successfully mapped to the five chromosomes of *A. thaliana*. On average, every base pair was

covered by 49 reads. Processing of these reads resulted in the identification of 3569 SNPs varying from the *A. thaliana* genome reference sequence. Of these SNPs, 1163 were not found in the sequenced *K/K;H/H* control plant. Analysis of these positions resulted in the identification of 66 homozygous SNPs causing non-synonymous mutations in the CDS of protein coding genes (Table S10).
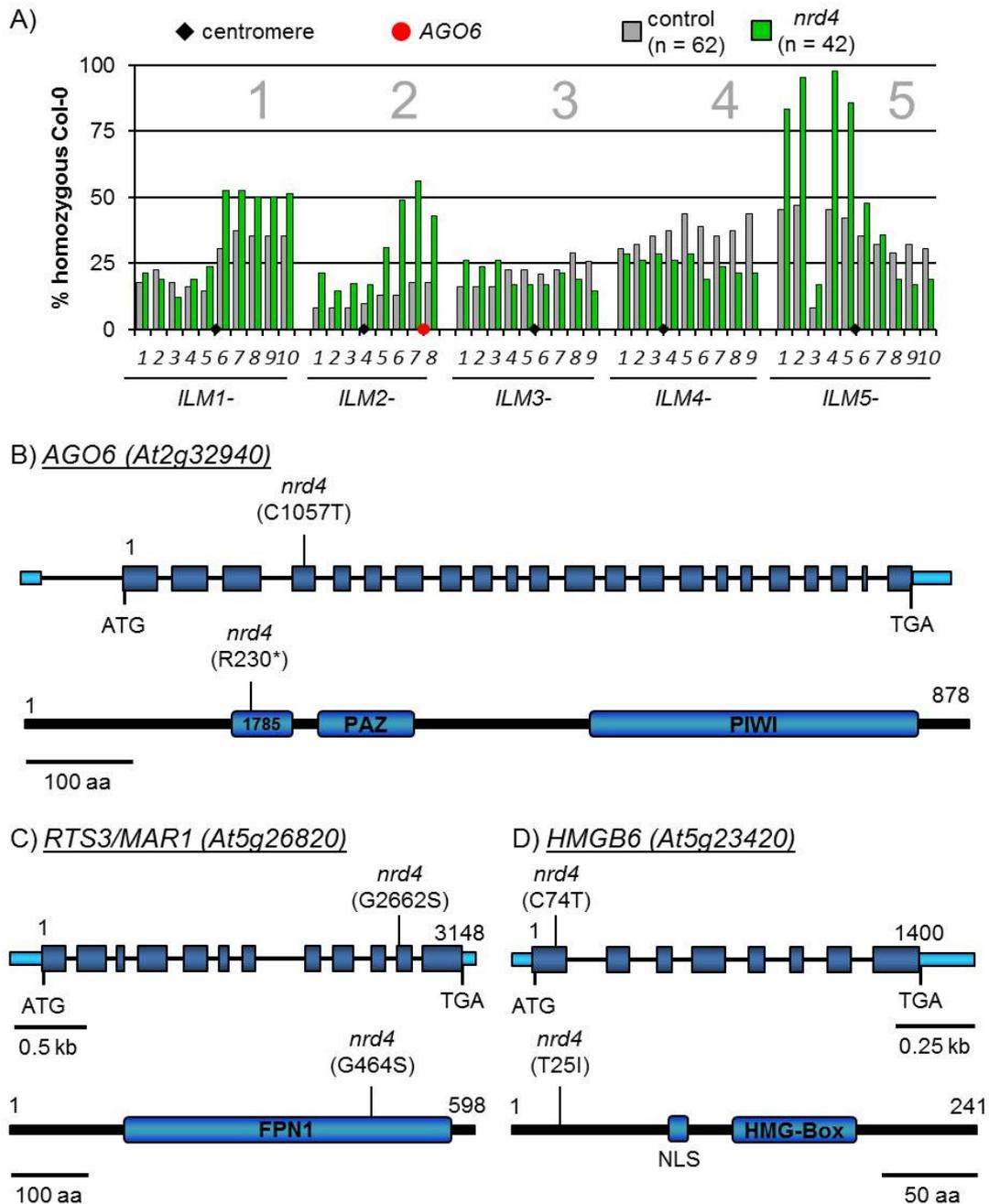
Figure 17: *nrd4*.

A) Marker allele incidences in a Kan[R] Hyg[R] M_3F_2 population derived from *nrd4* as determined Illumina GoldenGate assay. Displayed data include possible false positive individuals. For genotyping results using an extended mapping population of 117 individuals see Table S6. B – D) Gene and protein models indicating nucleotide and amino acid exchanges presumably contributing to the Kan[R] phenotype of *nrd4*. Numbering in gene models is according to the genomic sequence relative to the first bp of the START codon. Conserved proteins domains are indicated. DUF1785 (1785), PIWI/ ARGONAUT/ZWILLE (PAZ), Ferroportin1-like (FPN1), Nuclear localization signal (NLS), High Mobility Group box (HMG-Box).

Among these, 19 were located on chromosome 5 (Table S10). However, none of these mutations affected a gene locus previously described to be involved in RdDM or RdTGS.

Evaluation of the mutant loci for genes listed in the chromatin database (Gendler *et al.*, 2008) (www.chromdb.org) revealed that one mutation affects a gene locus coding for a chromatin-associated protein. This locus (*At5g23420*) codes for HMGB6, one of six HMGB proteins identified in *A. thaliana* (Grasser *et al.*, 2004; Merkle and Grasser, 2011). The identified C ➔ T mutation in the 1st exon at position 74 (relative to the A of the START codon) causes the exchange of threonine 25 for an isoleucine. The residue is part of the extended N-terminal region of HMGB6 which is unique among to the HMGB proteins of *A. thaliana*. Putative HMGB6 homologs of other plant species were identified by BLASTP searches using the full length HMGB6 as query. Protein sequence alignments of the identified homologs revealed no evolutionary conservation of the affected residue (data not shown).

A further mutation present in *nrd4* affected gene *At5g26820* coding for *MAR1/IREG3/RTS3*. Loss of the respective protein can cause resistance to aminoglycoside antibiotics such as kanamycin independent of the presence of transgenic detoxifying enzymes (Aufsatz *et al.*, 2009, Conte *et al.*, 2010).

Further analysis of identified SNPs lead to the identification of a C ➔ T mutation in the 4th exon of *AGO6* (*At2g32940*), a member of the AGO4 clade of *A. thaliana* AGO proteins that was previously described to act in RdDM and RdTGS (Zhang *et al.*, 2007; Havecker *et al.*, 2010; Eun *et al.*, 2011). The nonsense mutation at nucleotide 1057 causes a STOP codon at amino acid position 230 of AGO6, resulting in a predicted protein lacking all domains essential for AGO protein function.

To validate the identified *ago6* mutant allele as causative for the release of RdTGS in *nrd4*, complementation by introduction of an *AGO6* wild type ORF was initiated. However, transformation of M$_3$ individuals with the complementing T-DNA construct and subsequent analysis of positive transformants could not be completed within the time frame of the work for this thesis.

3.5.6 *NRD5* IS POTENTIALLY ALLELIC TO *DRD1*

Map based cloning

A mapping population of 113 plants, representing approximately 17% of initially 671 Kan^R Hyg^R M$_3$F$_2$ individuals derived from M$_3$ of line *2-11* was genotyped using the GoldenGate

assay to identify the location of the mutation causative for Kan^R Hyg^R phenotype. Genotyping resulted in the observation that all markers located at chromosome 2 were severely enriched for Col-0 alleles compared to the Hyg^R GUS$^+$ C$_3$F$_2$ control population, with marker *ILM2-5* showing the highest Col-0 allele incidence. However, a higher incidence of the homozygous Col-0 alleles was observed for markers located on the upper arm of chromosome 5 with marker *ILM5-2* and *ILM5-4* showing the highest incidence. Sequencing of *DRM1* and *DRM2* located between *ILM5-2* and *ILM5-4* did not lead to the identification of any mutation in the promoter or the CDS of these genes. Hence, whole genome sequencing was performed to identify the causative mutation in this candidate line.

Whole genome sequencing

NGS of one M$_3$ individual of *2-11* resulted in 66,747,201 unique reads of which 64,652,780 (96.9%) were properly mapped to sequences of the five *A. thaliana* chromosomes. On average a single base pair was covered by 47 reads. Processing of this reads lead to the identification of 1235 SNPs unique for *2-11*. Of these, 84 SNPs were found to cause homozygous non-synonymous mutations in the CDS of protein coding genes.

A missense mutation was identified in gene *At2g16390* which is located close to marker *ILM2-5.* The gene is coding for DRD1, a SWI/SNF2 chromatin remodeling factor known to be involved in RdTGS and RdDM. The G ➜ A mutation located in the 4th of the gene at position 2870 (relative to the 1st nucleotide of the START codon) causes the exchange of a glycine for an aspartate at position 683 of the protein. The homozygous nature of the mutation was verified by sequencing the PCR products of the respective genomic segment of *drd1* in five M$_3$ mutant plants. For estimation of a functional impairment caused by the amino acid exchange, protein sequence alignment of SWI2/SNF2 chromatin remodeling factors known and proposed to act in RdDM in *A. thaliana* as well as of putative DRD1 homologs of other plant species were performed. Results indicated a high conservation of the affected residue (Figure 18D), pointing to a role of this residue for functionality of the protein. Thus *2-11* is most likely allelic to *drd1* and represents a member *nrd5* a further complementation group. Due to time constrains, no verification by introduction of a DRD1 wild type ORF was carried out in the frame of this thesis. The putative *drd1* allele was termed *drd1-af*.
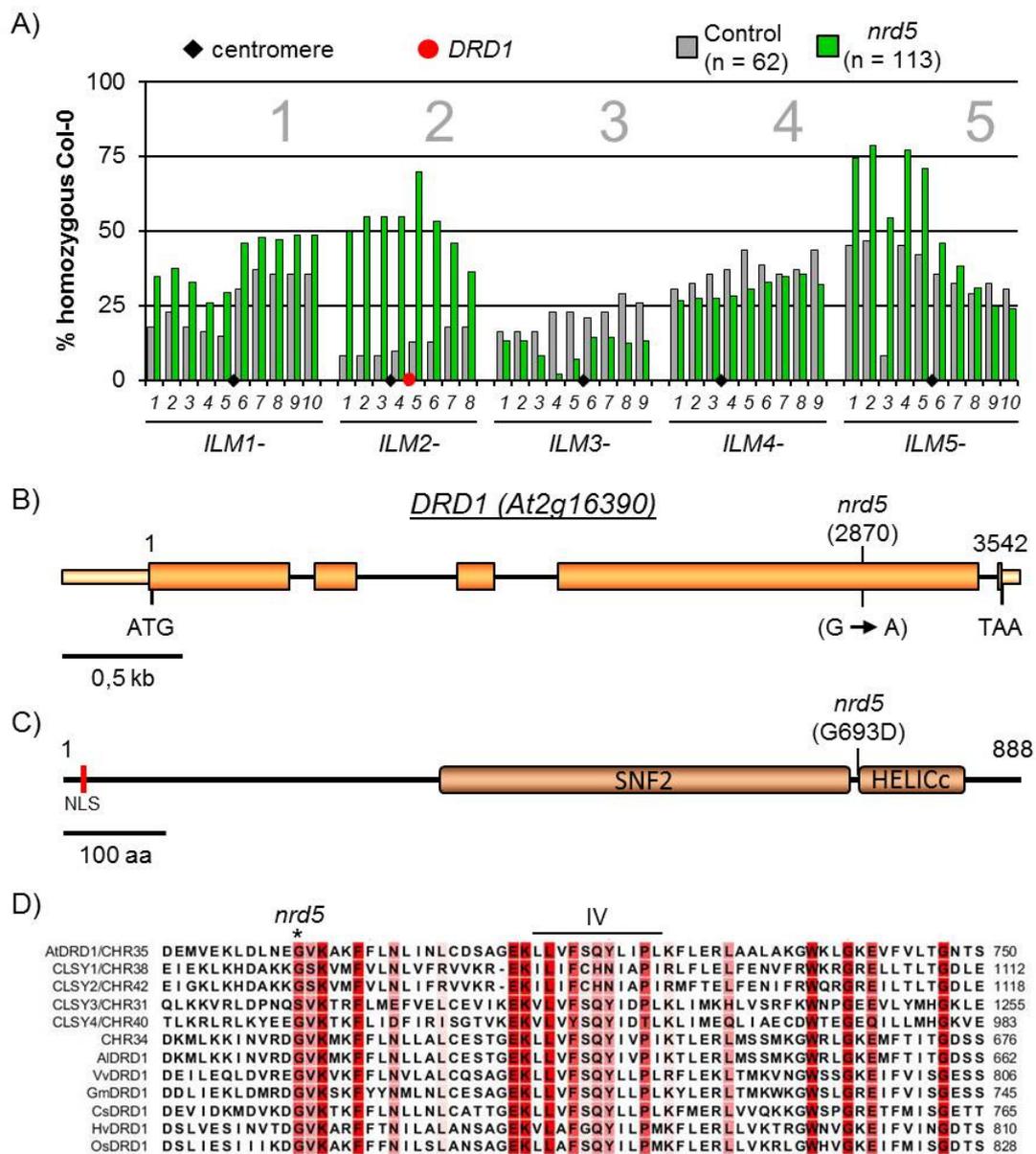
Figure 18: *nrd5*

A) Marker allele incidences in a Kan^R Hyg^R M₃F₂ population derived from *nrd5* as determined by GoldenGate assay. Values are based on results prior to removal of "false positive" plants. B) Gene model of *DRD1*. Position and nature of the nucleotide change identified by whole genome sequencing is depicted. Numbering is according to the genomic sequence relative to the first nucleotide of the START codon. C) Protein model of DRD1. Conserved domains are indicated. D) Fragment of a protein sequence alignment of SWI2/SNF2-like chromatin remodeling factors of *A. thaliana*, *Arabidopsis lyrata* (Al) *Vitis vinifera* (Vv), *Glycine max* (Gm), *Cucumis sativa* (Cs) *Hordeum vulgare* (Hv), Conserved residues are highlighted in red, The amino acid residue affected in *nrd5* is marked by an asterisk. For the identification of DRD1-like proteins, a full length putative DRD1 protein sequence was used. The ClustalW2 algorithm was used for alignment.

3.5.7 *NRD6* IS POTENTIALLY ALLELIC TO *NRPE1*

The mutations causative for the release of transcriptional silencing in mutants *9-20* and *11-12* were identified by NGS of one M₃ individual each. Sequencing of genomic DNA of *9-20* and *11-12* resulted in 57,551,632 and 43,326,794 unique reads, respectively. Of

72

these 53,350,327 (92.7%) and 35,668,197 (82.3%), respectively, were successfully mapped to the chromosomes of *A. thaliana* resulting in approximately 39-fold (*9-20*) and 26-fold (*11-12*) coverage of the genome.

In line *9-20*, 997 SNPs not found in the sequenced non-mutagenized *K/K;H/H* individual were identified. 79 of these caused homozygous, non-synonymous mutations in CDS of protein coding genes (Table S12). One of these mutations represents a G ➔ A exchange at position 4036 in the 9th exon of *At2g40030*. This gene encodes for NRPE1, the largest catalytic subunit of Pol V, which indispensable for RdDM (Pontier *et al.,* 2005; Kanno *et al.,* 2005). The identified mutation leads to the exchange of glycine 937 for an aspartate in the RNA polymerase domain of the protein. Genotyping of a mapping population of 17 Kan^R Hyg^R M_3F_2 plants derived from line *9-20* lead to the identification of a high incidence of the Col-0 allele for marker *ILM2-8* located physically close to *At2g40030 (NRPE1)*. This further supports the assumption that the identified *nrpe1* allele is causative for the release of RdTGS and RdDM in this line. Therefore, line *9-20* presumably contains with *nrd6-1* an allele of a 6th complementation group of the current screen.

To estimate the impact of the amino acid exchange on the functionality of the NRPE1 protein, putative homologous proteins in other plant species were identified by BLASTP search using full length *A. thaliana* NRPE1 protein sequence. Sequence alignments of the identified homologs revealed a complete conservation for the mutated residue in the analyzed proteins (Figure 19). In line *11-12*, 1069 unique SNPs that were not present in the sequenced non-mutagenized *K/K;H/H* individual were identified. 103 of these caused homologous, non-synonymous mutations in the CDS of protein coding genes (Table S13). As in line *9-20*, one mutation causes a single nucleotide exchange in *At2g40030* (*NRPE1*). The C ➔ T exchange at position 5495 in exon 13 of the gene causes a non-sense mutation leading to a premature STOP codon. The resulting protein is deficient of the C-terminal domain containing the reiterated GW/WG motives (AGO-hook domain) that was reported to be crucial for the interaction with AGO proteins. Therefore line *11-12* is as *nrd6-2* presumably the second member of the *nrd6* complementation group. Due to time constrains, no verification by transgenic introduction of a *NRPE1* wild type ORF or complementing crosses between *nrd6-1* and *nrd6-2* was carried out in the frame of this thesis. The putative loss-of-function alleles were termed *nrpe1-af1* and *nrpe1-af2*.
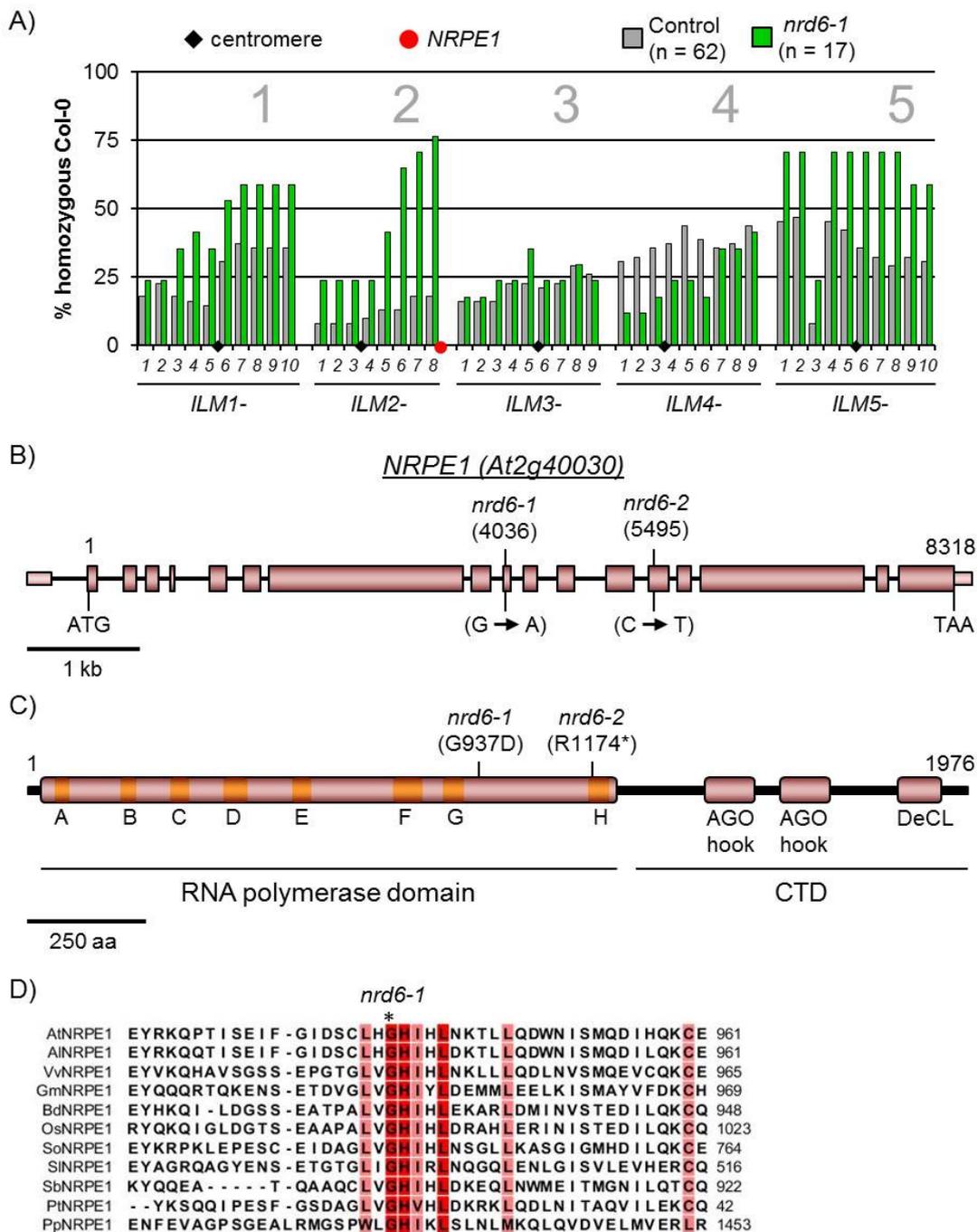
Figure 19: *nrd6*

A) Marker allele incidences in a Kan[R] Hyg[R] M_3F_2 population derived from *nrd6-1* as determined by Illumina GoldenGate assay. Depicted are results prior to removal of putative "false positive" individuals. B) Gen model of *NRPE1*. Positions and nature of identified nucleotide are indicated. Numbering is according to the genomic sequence relative to the first bp of the START codon. C) Protein model of NRPE1 showing positions of the predicted amino acid exchange and premature STOP codon, respectively. Conserved domains and motives are according to Pontier et al., (2005). D) Protein sequence alignments of *A. thaliana* NRPE1 (AtNRPE1) and putative NRPE1 homologs of *Arabidopsis lyrate* (Al), *Vitis vinifera* (Vv), *Glycine max* (Gm), *Brachipodium distachion* (Bd), *Oryza sativa* (Os), *Spinachia oleracea* (So), *Solanum lycopersicon* (Sl), *Populus trichocarpa* (Pt), *Physcomytrella patens* (Pp).

3.6 *DE NOVO* ASSEMBLY OF THE *ARABIDOPSIS THALIANA* GENOME

As described above, in several mapping approaches (*nrd4*, *nrd5* and *nrd6-1*) an unanticipated low incidence of homozygosity for Col-0 alleles was observed for markers *ILM5-3* and *C5-5609978* when genotyping F_2 individuals containing the *K* and *H* transgenes derived from *K/K*;*H/H* x L*er* crosses. This gave the impression of a "coldspot" of recombination in this area on chromosome 5. A notable exception was observed for *nrd3-1*, for which two peaks of high incidence of Col-0 alleles were observed, one in the same area of chromosome 5 and a second one on chromosome 2.

It is well known that, chromosomal rearrangements, like inversions, might occur upon *A. tumefaciens* mediated T-DNA transfer and have the potential to negatively affect recombination frequencies if present in heterozygous state (Andolfatto *et al.,* 2001; Kirkpatrick, 2010; Nacry *et al.,* 1998). To address the presumption that chromosomal rearrangements might cause the reduced recombination rate at marker *ILM5-3*, the sequence information obtained from whole genome sequencing was used to perform *de novo* assembly of the genome of the sequenced non-mutagenized *K/K*;*H/H* control plant as well as of M_3 *nrd3-2* and M_3 *nrd4* plants. Staggered assembly of the NGS reads lead to the formation of longer assembly units (contigs). The identification of two contigs that map to the same chromosome but show discordance in size and orientation between their 5' and 3' ends would indicate an inversion event (Korbel *et al.,* 2007; Le Scouarnec *et al.,* 2012). Mapping of contigs of *K/K*;*H/H* resulted in the identification of contigs 2597 and 2882 which map to the upper arm of chromosome 5. These contigs are 93 kb and 80 kb in length, respectively. However, their respective ends map approx. 800 kb apart. While the break point of contig 2597 was mapped to the intergenic region between loci *At5g13130* and *At5g13140*, the break point of contig 2882 maps between *At5g17200* and *At5g17210* (Figure 20). Pairwise alignment of the intergenic sequences led to the identification of two highly similar (96%) sequences of approx. 760 bp that are in sense / antisense orientation to each other. Notably, no such contigs were obtained during read assembly for mutant M_3 *nrd3-2*, making it conceivable that the respective rearranged version of chromosome 5 is segregating among line *K/K*;*H/H*-derived plants.

However, attempts to confirm this inversion by PCR using primers specific for flanking regions of the assumed break points were inconclusive (data not shown), leaving it possible that chromosome rearrangements are more complex.
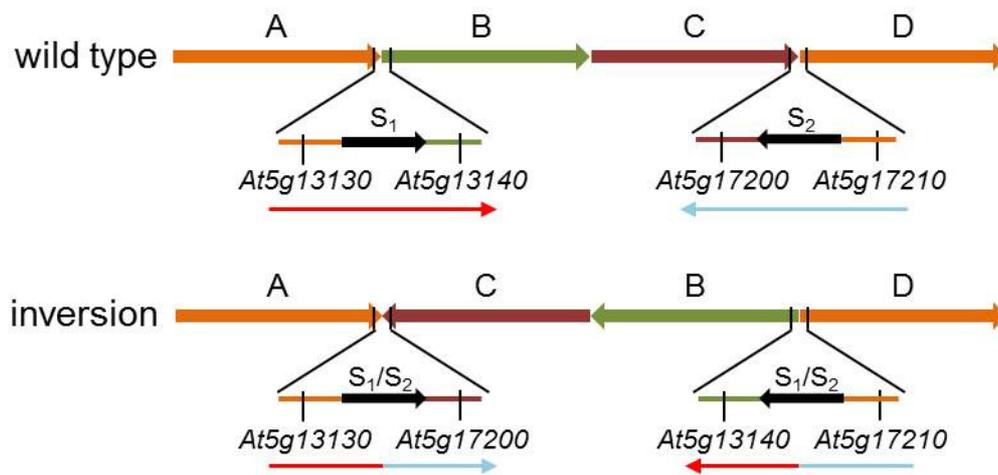
Figure 20: A putative inversion at chromosome 5

Order of loci in wild type plants according to the *A. thaliana* Col-0 genome reference sequence (upper panel) and in (some) transgenic *K/K;H/H* wild type plants as suggested by *de novo* assembly of NGS reads of the sequenced non-mutagenized *K/K;H/H* and M_3 *nrd4* individuals (lower panel). Locations and composition of highly similar sequences at break points are indicated (S1, S2, S1/S2).

4   DISCUSSION

To the present day, the forward genetic screen carried out in this thesis lead to the identification of ten independent *nrd* mutants which were assigned to six distinct complementation groups. Analysis of the DNA methylation at endogenous sequences revealed that the mutations in these mutants cause substantial reduction of methylation in non-CG context at known RdDM targets. The candidate gene loci in four of these *nrd* mutants were identified by map based cloning, whereas putative candidate gene loci in the remaining six *nrd* lines were identified by NGS. The mutations were found to affect components of the RdDM pathway previously isolated in other forward genetic screens (Kanno *et al.,* 2004; Kanno *et al.,* 2008; Eun *et al.,* 2012).

4.1   SILENCING OF THE EMPLOYED TRANSGENE SYSTEM DEPENDS ON THE RDDM PATHWAY

A transgene system targeting a *ProNOS-NPTII* reporter construct as employed in the current work was used in a previous genetic study of RdTGS of a transgene locus in *A. thaliana* (Aufsatz *et al.,* 2002a). No mutantion in RdDM factors were identified in this screen, but rather several *met1* and *hda6* alleles were isolated (Aufsatz *et al.,* 2002a; Aufsatz *et al.,* 2004). Based on the sole identification of factors involved in maintanance of CG methylation instead of RdDM components, it was suggested that the relatively high density of sites in the CG context in the *ProNOS* leads to a particular role of the maintenance pathway for silencing of *ProNOS* controled transgenes (Eun *et al.,* 2012). Interestingly, neither *met1* nor *hda6* were identified in our screen so far. Instead, all identified mutant alleles affect loci known to be involved in RdDM downstream of siRNA biogenesis. These loci were previously identified in other transgene systems involving different reporter gene promoters. This indicates that non-CG methylation is necessary for the stable silencing of the *ProNOS-NPTII* gene in the used *K* transgene. The most striking difference between both screens might be the copy number of th*K* transgene. While several complete and incomplete *K* insertion events were reported for the silencing system used by Aufsatz *et al.,* (2002a, 2002b), a single copy transgene particularly sensitive to RdDM was used for the screen performed in this thesis (Schubert *et al.,* 2004; Fischer *et al.,* 2008). Therefore, the different outcomes might be based on the different importance of diverse silencing mechanisms in dependence of the *ProNOS*

copy number rather than on sequence composition of the *ProNOS per se.* A further factor influencing the mechanism necessary for transcriptional silencing might be the local chromatin environment into which the transgene is inserted.

4.2 DEEPER SCREENING MIGHT RESULT IN THE IDENTIFICATION OF ADDITIONAL COMPLEMENTATION GROUPS

Due to the rather small number of isolated mutants releasing RdTGS and RdDM with three out of six complementation groups containing a single member so far, it is plausible that our screen is not yet saturated. This is further supported by the results of other forward genetic screens using similar transgene systems which managed to identify a wider spectrum of components of the RdDM pathway (Kanno *et al.,* 2008; Eun *et al.,* 2012 and references herein).

Thus further screening of our transgene system potentially result in the idenfication of new alleles of most of these and possibly even additional components. Designation of these mutations by map based cloning approaches might be challenged by the assumed translocation event affecting chromosome 2 and 5 as most of these loci are located at chromosome 2. Nevertheless, the availability of new NGS-based methods for mutation mapping might help to solve these shortcomings (Hartwig *et al.,* 2012).

4.3 MUTATION OF EVOLUTIONARY CONSERVED RESIDUES IN POL V SUBUNITS IMPAIR RDDM

In addition to the expected set of subunits for the common eukaryotic DNA-dependent RNA polymerases I, II and III, the annotation of the *A. thaliana* genome indicates additional genes which encode atypical largest and second-largest subunits for which plant specific functions were predicted (Arabidopsis genome initiative, 2000). It is now well established that these proteins represent the catalytic subunits of the two plant specific DNA-dependent polymerases Pol IV and Pol V, which are essential in RdTGS *via* RdDM (Herr *et al.,* 2005; Onodera *et al.,* 2005; Pontier *et al.,* 2005; Kanno *et al.,* 2005). Analysis of the subunit composition revealed that both complexes are constituted of 12 subunits, some of which shared with Pol II (Ream *et al.,* 2006; Luo and Hall, 2007; Law *et al.,* 2011; Eun *et al.,* 2012), supporting the assumption that Pol IV and Pol V evolved from Pol II (Ream *et al.,* 2006; Luo and Hall, 2007). In the course of this thesis mutations, in the subunit NRPD2a, common to Pol IV and V, and in the Pol V specific subunit NRPE1 were identified.

4.3.1 THE EFFECT OF MISSENSE MUTATIONS IN THE SECOND-LARGEST SUBUNIT

In *A. thaliana* two closely related genes putatively encoding a second-largest subunits of Pol IV and Pol V, *NRPD2a* and *NRPD2b*, were identified. Both genes were found to be transcribed (Pontier *et al.,* 2005). However, only *nrpd2a* mutants affect RdDM and only *NRPD2a* was identified to encode a functional subunit of Pol IV and Pol V (Herr *et al.,* 2005; Ream *et al.,* 2008). Therefore, *NRPD2b* is considered to be most likely a transcribed pseudogene (Pontier *et al.,* 2005).

In the course of the work for this thesis three *nrpd2a/nrpe2a* alleles were identified. All identified alleles contain missense mutations resulting in exchanges of amino acids which are conserved among the second largest subunits of DNA dependent RNA polymerases of *A. thaliana* and in case of *nrpd2a-54* (*nrd2-1*), in virtually all pro- and eukaryotic DNA-dependent RNA polymerases.

The DNA methylation in non-CG context was found to be severely reduced at *TARGET-ProNOS*, *AtSN1*, *MEA-ISR* in all isolated *nrdp2a* mutants. Furthermore, DNA methylation analysis of *AtMU1* performed in *nrpd2a-54* (*nrd2-1*) and *nrpd2a-55* (*nrd2-2*) also revealed a significant reduction in non-CG context methylation. The observed reduction of non-CG methylation at *AtSN1 and MEA-ISR* is in good agreement with results previously published for *nrpd2a* and *nrpd2a nrpd2b* double mutants (Kanno *et al.,* 2005a; Onodera *et al.,* 2005; Lopez *et al.,* 2011; X Zhang *et al.,* 2007b; Greenberg *et al.,* 2011).

The slight differences in non-CG methylation at *AtSN1* and *MEA-ISR* in *nrpd2a-54* and *nrpd2a-55* compared to data published for *nrpd2a* and *nrpd2a nrpd2b* mutants might be due to differences in the growth conditions or plant age between both studies. Furthermore, although it is considered as a pseudogene a slight influence of the *NRPD2b* locus cannot be excluded.

The elevated NPTII amounts observed in F_1 individuals from complementing crosses between *nrpd2a-54* and *nrpd2a-55*, but not in F_1 from backcrosses of the respective mutant lines to *K/K;H/H* wild type plants and the almost 100% incidence of the Kan^R phenotype in F_2 progeny of complementing crosses, but only about 25% of Kan^R plants in F_2 progeny of backcrosses confirmed the allelic nature of the mutations in these lines. In addition, no re-establishment of CHH methylation at *AtMU1* was observed in the analyzed F_1 individual of *nrpd2a-54 x nrpd2a-55* crosses compared to the respective M_4 individuals. Analysis of F_1 progeny of the backcrosses showed clearly more methylation in the CHH context, but wild type levels were not reached. These results have to be

regarded with caution, as half of the analyzed F_1 genome descended from the *K/K;H/H* wild type parent and therefore displayed wild type methylation levels from the beginning. A possible explanation would be that re-methylation at hypomethylated *AtMU1* in the presence of functional NRPD2a in F_1 from backcrosses is a slow process that takes longer than one generation. Such delayed methylation in RdDM has already been observed at transgenic targets (Fischer *et al.,* 2008). With hindsight, analysis of F_1 individuals of crosses between *nrpd2a* alleles and mutants affected in RdDM factors which are not part of the Pol IV and Pol V complexes, such as *drm2, rdr2* or *ago4* might have been a more informative option.

The mutation in *nrpd2a-54* affects the second glutamate residue of the GEME motif which is conserved in virtually all pro- and eukaryotic DNA dependent RNA polymerases known (Cromie *et al.,* 1999, Sidorenko *et al.,* 2010). Sidorenko *et al.,* (2010) reported an virtually identical mutation in the GEME motive in *MEDIATOR OF PARAMUTATION2* (MOP2), a homolog of *NRPD2a* in *Zea mays*. The *mop2-1* mutation was reported to affect paramutation at some investigated loci such as b1 in a dominant manner. However no indication was found for dominance of the *nrpd2a-54* mutation. In the mapping approach, 11% Kan^R Hyg^R M_3F_2 were counted in progeny derived from the *nrpd2a-54* cross to L*er*, which is well consistent with the 14% expected for a recessive mutation. Furthermore, as revealed by the persisting non-CG methylation at *AtMU1* in backcrosses of *nrpd2a-54* to *K/K;H/H*, no dominant effect is observed for this endogenous target as well. The assumption is further supported by an *nrpd2a* allele reported by Kanno *et al.,* (2005) in which a mutation causing an amino acid exchange in the GEME motif is recessive as well. However, as only a small number of targets were analyzed in this study and the dominant effect observed in *mop2-1* was reported to be locus specific, the existence of loci in *A. thaliana*, that are affected in a dominant fashion cannot be entirely excluded. Whole genome bisulfite sequencing of wild type plants and such heterozygous for the *nrpd2a/nrpe2a-54* allele would be of interest to clarify this issue. Loci, hypomethylated in plants that are heterozygous for the *nrpd2a/nrpe2a-54* would be likely candidate loci that are affected in a dominant way. Nevertheless, a locus-specific slow re-methylation of hypomethylated endogenous RdDM targets as discussed above for *AtMU1* would always offer an alternative explanation for a seemingly dominant mutation effect.

In contrast to *nrpd2a-54* and *nrpd2a-55*, the *nrpd2a* allele *nrpd2a-af* (*nrd2-3*) was not confirmed by complementation. However, the mutation of also a highly conserved residue makes it a very likely loss-of-function allele.

4.3.2 PUTATIVE LOSS-OF-FUNCTION ALLELES OF POL V LARGEST SUBUNIT

The most prominent difference between NRPE1, the largest subunit of Pol V, and the largest subunits of the other nuclear DNA-dependent RNA polymerases Pol I, Pol II, Pol III and Pol IV is an extended C-terminal domain (CTD) (Haag *et al.,* 2012). The Pol V CTD is characterized by a WG/GW-rich domain which is known to physically interact with AGO4, AGO6 and AGO9 as well as with SPT5L and is essential for the locus- specific recruitment of the above AGO proteins to chromatin (Li *et al.,* 2006; El-Shami *et al.,* 2007; Havecker *et al.,* 2010).

In the context of the work for this thesis, NGS of two mutant lines lead to the identification of mutations affecting NRPE1. The analysis of the DNA methylation status of endogenous RdDM targets in these lines revealed a severe decrease of non-CG methylation at all tested loci. The reduction of non-CG methylation in *AtSN1* determined by methylation-sensitive restriction cleavage using HaeIII was previously published for the alleles *nrpe1-1* (Pontier *et al.,* 2005), *nrpe1-11* (Lopez *et al.,* 2011) and an unnumbered *nrpe1* allele (Kanno *et al.,* 2005a). Furthermore, He *et al.,* (2007) provide *AtSN1* bisulfite sequencing data for the *nrpe1-11* allele. All these data are in good agreement with the results obtained for *nrpe1* candidates in this thesis work (He *et al.,* 2007; He *et al.,* 2009). Bisulfite sequencing data for *MEA-ISR* previously reported for *nrpe1-11* in accession Col-0 and an unnumbered *nrpe1* allele in accession C-24 reveal a similar loss of DNA methylation in non CG-context as observed for *nrpe1* alleles identified in our screen (He *et al.,* 2007; Liu *et al.,* 2011; Greenberg *et al.,* 2010).

Due to time constrains, complementation tests of these mutants were not carried out in the frame of this thesis. Furthermore, a possible effect of the mutation on the protein level was not assayed due to lack of appropriate antibodies. Therefore, the identified mutant alleles *nrpe1-af1* (*nrd6-1*) and *nrpe1-af2* (*nrd6-2*) still need to be considered as likely candidates. Nevertheless, the assumption that they are causative for the release of silencing and DNA methylation is supported by two further facts. First, genotyping of the *nrpe1-af1* mapping population reveal a high incidence of homozygosity for the Col-0 allele towards the end of the lower arm of chromosome 2, the location of the NRPE1 gene. Second, the amino acids affected by the mutations in the two alleles seem to be

conserved among NRPE1 homologs of multiple plant species, which hints to their functional importance. In contrast to *nrpe1-af1*, no genotyping of a mapping population was carried out for *nrpe1-af2*. The presence of a premature STOP codon in *nrpe1-af2* makes it a likely null allele.

4.4 Two new non-sense alleles in DOMAINS REARRANGED METHYLTRANSFERASE 2

The initial studies to address the roles of the members of the *DRM* gene family in DNA methylation in *A. thaliana* found *DRM2* but failed to detect *DRM1* mRNA in vegetative tissues (Cao and Jacobsen, 2002). Moreover, CHH maintenance as well as *de novo* methylation were found to be affected in *drm2* single and *drm1 drm2* double mutants, but not in *drm1* single mutant plants (Cao and Jacobsen, 2002; Cao and Jacobsen 2003). This led to the conclusion that *DRM2* encodes the only actively expressed DNA methyltransferase in *de novo* methylation and that *DRM1* is supposedly a pseudogene. However, the recently discovered egg cell-specific expression of *DRM1* and the reduced CHH methylation level at *MEA-ISR* during early embryo development in *drm1* mutants amend the earlier assumption, showing that *DRM1* is functional in a well-defined developmental stage (Jullien *et al.,* 2012).

Two *drm2* alleles were identified during the work for this thesis. Both mutations create premature STOP codons. In case of the *drm2-8* (*nrd3-1*) allele, which was confirmed by complementation, translation of the mRNA would most likely result in a non-functional protein lacking important motifs of the DMTase domain. As the loss of half of the catalytic motifs is accompanied by release of RdTGS, the loss of the entire DMTase domain in proteins derived from the second allele *drm2-af* (*nrd3-2*) certainly would cause a release of silencing, too.

4.5 A putative loss-of-function allele in *DRD1*

SWI2/SNF2 ATPases are chromatin associated enzymes that confer chromatin remodeling by repositioning of nucleosomes (Coe *et al.,* 1994). While binding to the targeted DNA is conferred in an ATP-independent manner, release of the DNA strand requires hydrolysis of the tri-phosphate (Vignali *et al.,* 2000). SWI2/SNF2 ATPases are characterized by two similar protein domains which contain seven conserved sequence motifs involved in binding and coordinated processing of ATP and DNA substrates (Smith and Peterson, 2005).

The SWI2/SNF2 chromatin remodeling factor-like proteins DDM1, CLASSY1 and DRD1 are known components of the of DNA methylation machinery in *A. thaliana*. While DDM1 mainly acts in co-operation with MET1 in maintenance of methylation in CG context, CLSY1 and DRD1 are components of the RdDM pathway. CLSY1 associates with Pol IV and RDR2. Hence, a function in early steps of RdDM is suggested (Smith *et al.,* 2007; Law *et al.,* 2011). By contrast, DRD1 acts in RdDM downstream of siRNA production. This assumption was originally made based on the observation that *drd1* mutants do not decrease the amount of 24 nt siRNAs derived from a constitutively transcribed *IR* of the α' promoter (Kanno *et al.,* 2004). In a follow up study, it was shown that DRD1 is not only involved in RdDM but also in the establishment of *de novo* methylation and has an active role in removal of CG methylation after the silencing signal is lost (Kanno *et al.,* 2005b).

In the work for this thesis, a missense mutation in *drd1* was identified by NGS. The mutation causes a G ➜ D exchange at position 693 of the protein. At all analyzed endogenous RdDM target sequences, a severe reduction of DNA methylation in the CHH context was observed comparable to the non-sense allele *drd1-6* (Kanno *et al.,* 2004; Chan *et al.,* 2006). While the data published for *MEA-ISR* are almost identical to my data, the results for *AtSN1* slightly differ in the way that substantially more non-CG methylation is retained in *drd1-af* (*nrd5-1*) than in *drd1-6*. This might be due to the different nature of the analyzed alleles or due to differences in experimental details. On the other hand, *drd1-af* exhibits the most severe reduction of DNA methylation at the *TARGET-ProNOS* of all identified mutant lines. This coincides with the highest amount of NPTII protein measured.

Due to time constrains, the causative role of the *drd1-af* mutation was not confirmed by complementation. However, several observations support the assumption that the mutation is essential for the release of TGS of the *NPTII* gene in this line. First, genotyping of the Kan[R] Hyg[R] M_3F_2 population revealed a high incidence of Col-0 homozygosity at markers close to the *DRD1* encoding gene locus. Moreover, the amount of *ProNOS* derived siRNAs was not found to be reduced but rather slightly increased in *drd1-af*. These observations are in agreement with those previously made for *drd1* null mutants in a similar transgene system (Kanno *et al.,* 2004). These authors reported a severe reduction of non-CG methylation at the α' promoter, which coincides with a strong release of GFP silencing. The observed release was stronger than that observed in *nrpd2a/nrpe2a* and *nrpe1* alleles isolated in the same screen (Kanno *et al.,* 2004).

Protein sequence alignments using the full length hydrolase domain (i.e. the SNF2 domain and DEXDc domain) of several SNF2-like proteins show that G693 is well conserved in the SNF2-like proteins that are known and/or implied to be involved RdDM in *A. thaliana*, pointing to its functional importance. G693 is in vicinity to the conserved motive IV which is involved in DNA binding. Its exchange might induce conformational or environmental changes that might impair this interaction. A further possibility is the impairment of viable protein-protein interactions, e.g. to NRPE1, DMS3 or RDM1.

4.6   AGO6 MIGHT BE NECESSARY FOR SILENCING IN TRANSGENE SYSTEMS

One of the identified mutants lines, *nrd4-1*, clearly differs from the remaining ones in respect to release of *NPTII* silencing and *TARGET-ProNOS* methylation. In this line only a partial release of the *NPTII* silencing was observed. Furthermore, substantial amounts of CHH and CHG context remained at the *TARGET-ProNOS*.

The analysis of the NGS data from a *nrd4-1* plant lead to the identification of a nonsense mutation in locus *At2g32940,* which encodes the RdDM factor AGO6. NULL alleles of *AGO6* were previously isolated in two independent forward genetic screens (Zheng *et al.,* 2007; Eun *et al.,* 2011). Both studies report an incomplete release of the expression of transgenic reporter genes in these mutants, which comes along with notable levels of CHH methylation remaining at the transgenic promoters. The premature STOP in *ago6-af* (*nrd4-1*) causes the truncation of the protein after 320 amino acids. The resulting protein would lack the PAZ and PIWI domains which are of vital importance for functionality of AGO proteins (Vaucheret, 2008). Therefore, it is valid to assume that the identified *ago6-af* is a null allele. Nevertheless, due to missing complementation it still has to be considered as a likely candidate.

In contrast to the *TARGET-ProNOS*, which retained substantial non-CG context methylation, the analyzed endogenous RdDM targets *AtSN1*, *MEA-ISR* and *AtMU1* exhibit DNA methylation levels which resemble those of other RdDM mutants identified in this thesis. Further, the results obtained for *AtSN1* and *MEA-ISR* are in line with those published for allele *ago6-1*. However, a rather mild effect on the CHH methylation in *AtSN1* was reported for the T-DNA insertion allele *ago6-2* (Zheng *et al.,* 2007; Havecker *et al.,* 2010). These deviant results may be due to differences between the *A. thaliana* accessions in which the analyzed alleles were obtained, the analyzed plant tissues or the

number of generation for which the analyzed plants had been homozygous for the respective *ago6* alleles.

The differences between the methylation levels of transgenic and endogenous sequences might be based on the different sources of dsRNA. While the synthesis of initial endogenous dsRNA depends on the activity of Pol IV and RDR2, dsRNA formation in the transgene system involving transcribed IRs depends on the activity of Pol II. Furthermore, the massive and constitutive transcription of the *ProNOS-IR* and comparably high amounts of *ProNOS* siRNAs in the transgene system might contribute to the observed effect.

AGO6 as well as AGO4, AGO8 and AGO9 belong to a clade of *A.thaliana* AGO proteins that bind 24 nt siRNAs and confer chromatin modifications. Within this clade, AGO4, AGO8 and AGO9 are closely related, while AGO6 shows higher sequence variation (Vaucheret, 2008). AGO4 is ubiquitinary expressed (Zilberman *et al.,* 2003; Havecker *et al.,* 2010; Eun *et al.,* 2011), expression of AGO6 is restricted to apical meristems and the connecting vascular tissue (Havecker *et al.,* 2010; Eun *et al.,* 2011). AGO9 expression seems to be limited to ovules, anthers and the seed coat (Havecker *et al.,* 2010; Olmedo-Monfil *et al.,* 2010, Calaroco and Martienssen, 2011). AGO8 is considered to be a pseudogene (Vaucheret, 2008). It is assumed, that AGO4 clade proteins act at least partially redundant (Havecker *et al.,* 2010). Based on this assumption the observed decrease of CHH methylation levels at *AtSN1, MEA-ISR* and *AtMU1* is striking as it suggests that the expression of AGO4 in leaf tissue is not sufficient to complement the lack of AGO6 in *ago6* mutants at these loci. This points to a possible functional non-redundancy between AGO6 and AGO4. The inability of a *ProAGO4:AGO6* construct to complement an *ago4-3* mutant further supports this idea (Havecker *et al.,* 2010). Possibly, the activity of AGO6 in the apical meristems is necessary to establish certain prerequisites, e.g. basal levels of DNA methylation which facilitate efficient recruitment of AGO4 to certain loci. However, this hypothesis will need experimental proof.

4.7 SECOND SITE MUTATION MIGHT INFLUENCE KAN[R] IN *AGO6*

Although the identified *ago6* mutation represents a very likely candidate to be responsible for the observed silencing release, results from genotyping of the Kan[R] Hyg[R] M_3F_2 mapping population are somewhat inconsistent with this assumption. Indeed, they revealed a markedly increase of homozygosity for Col-0 alleles at markers physically close to the *AGO6* locus. However, a much higher incidence of homozygosity for the Col-

0 allele was found for markers at the upper arm of chromosome 5. Together with the observed incidence of approximately 7.5 % for the Kan^R Hyg^R phenotype in the M_3F_2 generation for the *nrd4-1*, this observation suggests a second mutation in this line that contributes to the Kan^R phenotype. Two mutated loci that might contribute to the Kan^R Hyg^R phenotype were identified by NGS on the upper arm of chromosome 5.

4.7.1 A HMG-box containing protein might affect transcriptional silencing

One of these mutations affects *HMGB6*, the gene encoding a nuclear chromatin-associated protein of unknown function, which is able to bind to DNA *via* its HMG-box domain *in vitro* (Grasser *et al.,* 2004). Due to its association with chromatin, involvement of HMGB6 in epigenetic processes, like RdDM, is plausible. The identified mutation would cause the exchange of threonine 25 (T25) for an isoleucine in the N-terminal domain. Protein sequence alignments with sequences of putative HMGB6 homologs identified by BLASTP searches revealed no evolutionary conservation of the residue, leaving is significance open.

4.7.2 RdDM independent mutation might contribute to the Kan^R phenotype of ago6.

In addition to the identified putative *ago6* and *hmgb6* alleles, a mutation in locus *At5g26820* was identified in *nrd4-1*. This locus encodes RTS3/MAR1, a chloroplast protein with similarity to the iron regulated transporters IREG1 and IREG2. Two previous studies reported that loss-of-function *rts3/mar1* allele can confer increased resistance to kanamycin, most likely due to the impaired antibiotics uptake (Aufsatz *et al.,* 2009; Conte *et al.,* 2009).

The G464S mutation identified in *nrd4-1*, affects a residue which is invariant between RTS3/MAR1 homologs in *A. thaliana, O. sativa* and *V. vinifera* (Conte *et al.,* 2009). Therefore, its mutation might cause a non-functional protein that contributes to the Kan^R phenotype in a RdDM independent manner. This would explain the high incidence of homozygosity for Col-0 alleles at the upper arm of chromosome 5 observed in Kan^R Hyg^R M_3F_2 individuals. To test if the rts3/*mar1* allele influences the segregation of the Kan^R Hyg^R phenotype, genotyping of mapping populations obtained from homozygous *rts3/mar1* single mutants would be necessary.

4.8 A POINT MUTATION IN THE XH DOMAIN OF IDN2 CAUSES RELEASE OF RdTGS

In the genome of *A. thaliana* 14 gene loci encoding proteins with sequence homology to SUPRESSOR OF SILENCING (SGS3) and its rice homolog gene X1 that fell into three clades were identified (Qin *et al.,* 2009; Xie *et al.,* 2012). SGS3-like proteins are characterized by various combinations of the XS domain, the XH domain, a coiled coil domain (CC) and a zinc finger domain (ZnF) (Bateman, 2002). IDN2 harbors all of these domains. *In vitro* studies showed the ability of IDN2 to bind dsRNA with 2 nt 5'-overhangs and with blunt ends via its XS domain. It is assumed that this preference is conferred by special structural elements present in the RNA recognition motif of the XS domain. The vital role of IDN2 in RdDM was identified in a screen for T-DNA insertion mutants incapable to methylate a FWA transgene *de novo*. Additional *idn2* alleles were identified in independent forward genetic screens for suppressors of RdDM (Ausin *et al.,* 2009; Zheng *et al.,* 2010; Xie *et al.,* 2012; Lorcovice *et al.,* 2012; Finke *et al.,* 2012b).

In the progress of work for this thesis, an *idn2* allele was identified in one of the silencing suppressor mutants and was termed *idn2-8* (*nrd1*). The CHH context methylation levels at the *TARGET-ProNOS* were found severely decrease in *idn2-8* and resembled those of other silencing suppressor mutants obtained in this screen. The reduction of *TARGET-ProNOS* methylation was accompanied by a markedly increase of NPTII protein levels which resembled those observed of the identified *nrpd2a/nrpe2a*, *drm2* and *nrpe1* alleles, but was somewhat lower than those of the *drd1* allele and the non-silenced wild type *K/K;-/-* control. The persistence of reduced NPTII expression might be explained by the residual CG and CHG context methylation at the *TARGET-ProNOS*, which might contributes to the silencing effect and is absent in non-silenced *K/K;-/-* plants. The nearly complete loss of CHH context methylation at *AtSN1*, CHH and CHG methylation at *MEA-ISR* and the markedly reduction of CHH methylation at *AtMU1* in identified *idn2-8* allele is similar to the results previously reported for the null alleles *idn2-1* and *idn2-5/rdm12-1*, respectively (Ausin *et al.,* 2009; Zhang *et al.,* 2012; Ausin *et al.,* 2012). Therefore, it was suggest that the identified point mutation in the XH domain renders *idn2-8* into an effective null allele. This was confirmed onto introduction of an IDN2 wild type ORF into *idn2-8* which complements the mutant phenotype.

As *idn2-8* mRNA level were not reduced compared to wild type, the mutant effect is not due to transcript degradation but most likely due to post-translational issues. IDN2 is supposed to act in higher order complexes consisting of IDN2 homodimers or

heterodimers formed with closely related IDP/FDM/IDNL proteins (Zhang *et al.,* 2012; Ausin *et al.,* 2012; Xie *et al.,* 2012, Wierzbicki *et al.,* 2012). It was shown that the homodimerization is necessary for the activity of IDN2 in RdDM and that it depends on the presence of the CC domain, whereas the XH domain is necessary for the formation of heterodimers (Ausin *et al.,* 2012; Zheng *et al.,* 2012; Zhu *et al.,* 2012). In *idn2-8* the formation of these higher order complexes might be compromised due to the structural alterations in the XH domain.

4.9   IDN2 ACTS DOWNSTREAM OF SIRNA FORMATION

As the RNA species fitting the binding characteristics of IDN2 occur more than once in the RdDM pathway, the position of *IDN2* in the RdDM pathway was a matter of debate in the initial publications. Based on analogy to the cooperation of SGS3 and RDR6 to generate dsRNA in PTGS, IDN2 has been proposed to cooperate in the RDR2-dependent production of dsRNA from a Pol IV-generated single stranded transcript (Zhang *et al.,* 2010). In an alternative model, IDN2 is involved in stabilizing siRNA-p5-RNA duplexes in the process of guiding DRM2-mediated DNA methylation (Ausin *et al.,* 2009). As formation of *ProNOS* 24nt siRNAs is conferred by Pol II-dependent transcription of an inverted repeat, the RdDM in this system works according to a linear rather than circular pathway (Figure 21). It was previously shown that generation of primary siRNAs and the establishment of RdDM are independent of Pol IV and RDR2. Like in the isolated *nrpd2a-55* mutant, siRNA formation was also not compromised in *idn2-8*. However, the CHH context methylation at the *TARGET-ProNOS* was severely affected by the *idn2-8* mutation. Thus, *idn2-8* needs to act downstream of the siRNA formation in RdDM (Finke *et al.,* 2012b). This assumption is in agreement with an only partial reduction of 24 siRNAs, derived from endogenous RdDM targets such as *AtSN1 MEA-ISR* and *AtMU1* observed in other *idn2/rdm12/dms10* mutant, while a marked impact on CHH and CHG context methylation was observed (Ausin *et al.,* 2009; Zheng *et al.,* 2010; Zhang *et al.,* 2012).

The assumption is further supported by *in vitro* experiment with the IDN2-interacting protein FDM1/IDP1/IDNL1 which showed physical interaction with AGO4 but not with RDR2 and by the reported binding of IDN2 to p5-RNAs(Zhu *et al.,* 2012; Xie *et al.,* 2012).
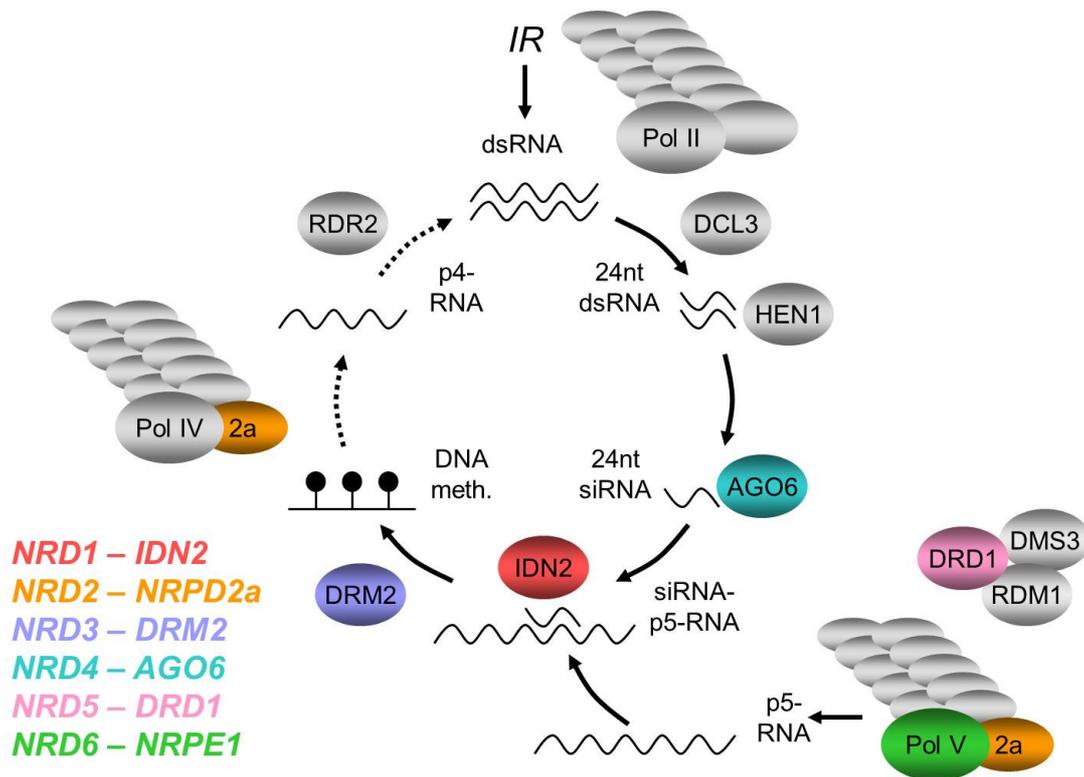
Figure 21: Genetic model of RdTGS in the transgene system.

RdTGS in the transgene system is independent of the Pol IV part of the pathway (dotted arrow) as transcription of an inverted repeat (IR) by Pol II leadto the formation of the dsRNA substrate of DCL3. Therefore RdTGS in this "shortcut" –pathway (solid arrows) depends on factors downstream of the siRNA formation. Factors identified in our screen with at least one allele are highlighted.

4.10   DNA METHYLATION AT *IGN* SEQUENCES AND *BASHO210*

In the progress of the work for this thesis, several less well characterized endogenous sequences were analyzed in respect to their DNA methylation status by methylation-sensitive restriction cleavage and / or bisulfite sequencing in a subset of the isolated mutants. The analyzed sequences were *IGN5*, *IGN23*, *IGN25* and *BASHO210*.

4.10.1 DNA METHYLATION AT *IGN5, IGN23* AND *IGN25*

The evaluation of CHH context methylation at *IGN5*, *IGN23*, *IGN25* in transgenic, but otherwise wild type plants with the help of methylation-sensitive restriction confirmed previously published results (Wierzbicki *et al.,* 2009; Rowley *et al.,* 2010; Lorkovic *et al.,* 2012). Bisulfite sequencing of *IGN5* in transgenic *K/K*;*H/H* wild type plants revealed a high amount of CG, CHG and CHH context methylation at this sequence. The obtained values for CG and CHG context methylation at this locus are in line with results published by others (Greenberg *et al.,* 2011). However, according to Greenberg *et al.,* (2010), less than 10 % of CHH context cytosines are methylated in wild type *A. thaliana* accession Col-0 plants. By contrast, approximately 30 % of cytosines in CHH context

were found to be methylated in *K/K*;*H/H* individuals. As both studies do not differ in analyzed accession or the analyzed sequence, the most likely explanation for these differences are variations in plant age, analyzed tissue or growth conditions.

Bisulfite sequencing of locus *IGN23* in wild type tissue indicated substantial DNA methylation in all sequence contexts in *K/K*;*H/H* individuals also outside of the two HaeIII restriction sites. These data are contradicted by a recent report, which could not detect any CHH context methylation at the *IGN23* locus in Col-0 plants (Wierzbicki *et al.,* 2012). In the study by Wierzbicki et al., (2012) NGS of bisulfite treated genomic DNA was used to assay genome wide DNA methylation. The criteria used to define methylated and unmethylated sequences might be causative for the consideration that *IGN23* is unmethylated.

The results for *IGN5* methylation in the assayed alleles of *idn2, drm2, drd1 and nrpe1* are in agreement with previously reported data for other mutant alleles of these loci (Lorkovic *et al.,* 2012; Wierzbicki *et al.,* 2009; Rowley *et al.,* 2010; Greenberg *et al.,* 2011). The effect of *nrpd2a* mutations on the DNA methylation at *IGN5* was not assayed previously. However, the observed decrease of CHH context methylation in *nrpd2a-54* is in agreement with the function of the protein in the Pol V complex. In contrast to these mutants, almost wild type level of DNA methylation at *IGN5* was observed in *ago6-af*. This is a striking difference compared to the other endogenous targets, in which the reduction in CHH methylation resembles the one in other mutants. This hints to a functional redundant protein, able to establish DNA methylation at this site in absence of *AGO6*.

4.10.2 THE RDDM PATHWAY ACTS AT *BASHO210*

In addition to the established RdDM targets, *BASHO210*, which was previously not described to be targeted by RdDM, was analyzed *via* bisulfite sequencing.. Sequencing of the 5'-part of the sense strand of *BASHO210* in wild type plants revealed substantial DNA methylation in all sequence contexts. While the observed levels of CG context methylation are in rather good agreement with published whole genome bisulfite sequencing data, the high levels of CHG and CHH context methylation were not reported by others (Stroud *et al.,* 2012). However, the high amounts of CHH context methylation which are especially enriched around the CTRR motif of *BASHO210* are in line with the high abundance of 24 nt siRNA which correspond to this locus (Lister *et al.,* 2008). The severely reduced CHH methylation levels in *idn2-8*, *nrpd2a/nrpe2a-54*, *drm2-8* and *ago6-*

af confirm the assumption that CHH context methylation at *BASHO210* depends on the RdDM pathway.

5   CONCLUSIONS

The results obtained in this work lead to the main conclusions:

1. The two component transgene system used in this thesis is suitable to identifiy components of the RdDM pathway in a forward genetic screen.

2. RdDM in the used transgene system is independent of Pol IV activity but requires a functional Pol V complex.

3. The member of the AGO4 clade of A. thaliana AGO proteins might have only limited functional redundancy and AGO6 might be necessary for efficient DNA methylation by AGO4..

4. IDN2 acts downstream of the generation of siRNAs.

5. The rolling circle DNA transposon *BASHO210* is targeted by the RdDM pathway.

6 OUTLOOK

1. To test whether the mutations in *nrpd2a-af*, *drm2-af, ago6-af*, *drd1-af*, *nrpe1-af1* and *nrpe1-af2* are causative for the release of silencing, crosses with other loss-of-function alleles or complementation with wild type ORFs has to be carried out.

2. To address the question if mutations in the GEME motif of NRPD2a might affect some endogenous loci in a dominant way, comparison of whole genome bisulfite sequencing analysis of plants heterozygous for *nrpd2a-54* and *K/K*;*H/H* wild type individuals would be suitable.

3. To test if the meristem specific activity of AGO6 is a prerequisite for efficient AGO4 activity, DNA methylation at endogenous targets in *ago6* mutants expressing a *ProAGO6:AGO4* construct should be carried out. Another important test for the involvement of AGO4 would be introgression of an *ago4* mutation into the transgene system used.

4. The fact that similar transgene systems identified a large number of RdDM factors and are supposedly saturated suggests that the potential of the used silencing systems to identify new RdDM factors almost reached its limits. To identify new, potentially functional redundant RdDM factors, affinity purification experiments coupled with mass spectrometric analysis might be carried out.

5. While there is meanwhile good insight into the genetic requirements for the establishment of RdDM, the knowledge concerning the mechanisms that prevent sequences being targeted by RdDM are rather poorly understood. A forward genetic screen *via* "enhancers of silencing" mutants would allow an approach to this.

7   SUMMARY

The aim of this work was the genetic identification and functional characterization of factors involved in RNA-directed transcriptional gene silencing (RdTGS) *via* induced loss-of-function mutants in *A. thaliana*. To isolate these, a transgene system was used in which two copies of the *NOPALINE SYNTHASE* promoter (*ProNOS*), arranged as an inverted repeat are constitutively transcribed. The resulting RNA mediates *in trans* hypermethylation of an unlinked *ProNOS* copy and hence the stable transcriptional repression of a *NPTII* reporter gene, resulting in a kanamycin-sensitive phenotype in wild type individuals.

Batches of M_2 seeds from ethyl methanesulfonate (EMS) mutagenesis of the transgenic line were screened for individuals displaying re-established kanamycin resistance (Kan^R). To confirm the release of *NPTII* silencing, NPTII protein levels in Kan^R mutants were analyzed using ELISA. Furthermore, DNA methylation patterns of the transgenic *ProNOS* as well as of several endogenous sequences were analyzed.

The genes affected by mutations causing a release of *NPTII silencing* in the isolated <u>n</u>o <u>r</u>na-<u>d</u>irected transcriptional silencing (*nrd*) mutants were determined by map based cloning and next generation sequencing. The *nrd* mutants could be grouped into six putative complementation groups: *nrd1/idn2*, *nrd2/nrpd2a*, *nrd3/drm2*, *nrd4/ago6*, *nrd5/drd1* and *nrd6/nrpe1.*

Nrd1 and *nrd3* were confirmed to be *idn2* and *drm2* loss-of-function alleles, respectively, by introduction of transgenes carrying wild type alleles of the respective candidate genes which resulted in reestablishment of RdTGS of the NPTII gene. The identity of *nrd2* as *nrpd2a* was checked in crosses between independent lines *nrd2-1* and *nrd2-2* both carrying mutations in *NRPD2a*. Unchanged NPTII levels and persistent DNA hypomethylation of an endogenous target in F_1 progeny confirmed that the defective *nrpd2a* alleles are indeed causing the release of silencing.

Furthermore, based on Northern analysis of *ProNOS*-derived siRNAs, it was concluded that IDN2 most likely acts downstream of siRNA formation in RdTGS.

8 ZUSAMMENFASSUNG

Das Ziel dieser Arbeit war die Identifizierung und funktionelle Charakterisierung von Komponenten die am RNA-vermittelten transkriptionellen Silencing und an der RNA-vermittelten DNA-Methylierung im Modelorganismus *A. thaliana* beteiligt sind. Dazu wurde ein Transgensystem verwendet in dem zwei, als *inverted repeat* angeordnete Kopien des *NOPALINSYNTHASE* Promoters (*ProNOS*) konstitutiv transkribiert werden. Die resultierende RNA vermittelt die Hypermethylierung einer weiteren, ungekoppelten Kopie des *ProNOS* in trans. Dadurch wird die Expression des Kanamcinresistenszgens *NEOMYCIN PHOSPHOTRANSFERASE II* (*NPTII*) unterdrückt. Die resultierenden Wildtyppflanzen sind trotz des *NPTII* Gens empfindlich gegenüber Kanamycin.

Die M_2 Generation von mittels EMS mutagenisierter transgener Samen, wurde auf Individuen getestet, die Resistenz gegenüber Kanamycin aufwiesen. Um die Aufhebung des transkriptionellen Silencings in den Kanamycin-resistenten Linien zu bestätigen, wurde die Menge des NPTII Proteins mittels ELISA bestimmt. Darüber hinaus wurden die DNA Methylierungsmuster des transgenen *ProNOS* sowie mehrerer endogener Sequenzen analysiert.

Die mutierten Gene, die für die Aufhebung des NPTII-Silencings in den isolierten *no rna-directed transcriptional silencing* (*nrd*) Mutanten verantwortlich waren, wurden mittels *Map-based Cloning* und *Next Generation Sequencing* identifziert. Die *nrd* Mutanten konnten so in die sechs putative Komplementationsgruppen *nrd1/idn2*, *nrd2/nrpd2a*, *nrd3/drm2*, *nrd4/ago6*, *nrd5/drd1* and *nrd6/nrpe1* eingeordnet werden.

Durch Einbringen eines Wildtyp Transgens des jeweiligen Kandidatengenes das zur Reetablierung des NPTII Silencings führte, wurde bestätigt, dass *nrd1* und *nrd3-1* loss-of-function Allele von *IDN2* bzw. *DRM2* sind, die für die Aufhebung des Silencings verantwortlich waren. Die unverändert hohen Mengen an NPTII Protein sowie andauernde DNA Hypomethylierung endogener Sequenzen in F_1 Individuen aus Kreuzung von *nrd2-1* mit *nrd2-2*, bestätigten dass es sich bei diesen Mutanten um nicht funktionelle Allele von *nrpd2a* handelt.

Da *nrd1/idn2* keine merklichen Verringerung der vom *ProNOS-IR* abgeleiteten 24 nt siRNAs aufwies, wurde geschlossen, dass IDN2 nach der siRNA Synthese von Bedeutung im RdDM ist.

9 REFERENCES

Abe M, Yoshikawa T, Nosaka M **(2010)** WAVY LEAFS1, an ortholog of Arabidopsis HEN1 regulates shoot development by maintaining MicroRNA and trans acting small interfering RNA accumulation in rice. *Plant Physiol.* **154**:1335-1346.

Agorio A, Vera P. **(2007)** ARGONAUTE4 is required for resistance to Pseudomonas syringae in *Arabidopsis. Plant Cell* **19**:3778-3790.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Chandler VL. **(2006)** An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**:295-298.

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

Andolfatto P, Depaulis F, Navarro A. **(2001)** Inversion polymorphisms and nucleotide variability in *Drosophila. Genet Res.* **77**: 1–8.

Arabidopsis Genome Initiative. **(2000)** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* **408**:796-815.

Arciga-Reyes L, Wootton L, Kieffer M, Davies B. **(2006)** UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in *Arabidopsis. Plant J.* **47**:480-489.

Aufsatz W, Mette MF, van der Winden J, Matzke AJ, Matzke M. **(2002a)** RNA-directed DNA methylation in *Arabidopsis. Proc Natl Acad Sci USA.* **99** Suppl 4:16499-506.

Aufsatz W, Mette MF, van der Winden J, Matzke MA, Matzke AJM. **(2002b)** HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double stranded RNA. *EMBO J.* **21**:6832-6841.

Aufsatz W, Mette MF, Matzke AJ, Matzke M. **(2004)** The role of MET1 in RNA-directed *de novo* and maintenance methylation of CG dinucleotides. *Plant Mol Biol.* **54**:793-804.

Aufsatz W, Nehlin L, Voronin V, Schmidt A, Matzke AJM, Matzke MA. **(2009)** A novel strategy for obtaining kanamycine resistance in *Arabidopsis thaliana* by silencing an endogenous gene encoding a putative chloroplast transporter. *Biotechnol J.* **4**:224-229.

Ausin I, Mockler TC, Chory J, Jacobsen SE. **(2009)** IDN1 and IDN2 are required for de novo methylation in *Arabidopsis thaliana. Nature Struct Mol Biol* **16**:1325-1327.

Ausin I, Greenberg MV, Simanshu DK, Hale CJ, Vashisht AA, Simon SA, Lee TF, Feng S, Meyers BC, Española SD, Wohlschlegel JA, Patel DJ, Jacobsen SE. **(2012)** INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proc Natl Acad Sci USA* **109**:8374-8381.

Banks JA, Masson P, Fedoroff N. **(1988)** Molecular mechanisms in the developmental regulation of the maize Suppressor mutator transposable element. *Genes Dev.* **2**:1364-1380.

Bannister AJ, Kouzarides T. **(2011)** Regulation of chromatin by histone modifications. *Cell Res.* **21**:381-395.

References

Bateman A. (2002) The SGS3 protein involved in PTGS finds a family. *BMC Bioinformatics*. **3**:21.

Bäurle I, Smith L, Baulcombe DC, Dean C. **(2007)** Widespread role of flowering time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science* **318**:109-112.

Bäurle I, Dean C. **(2008)** Differential interactions of the Autonomous Pathway RRM proteins and chromatin regulators in the silencing of Arabidopsis targets. *PLoS One* **3**:e2733.

Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, Woodcock CL. **(1998)** Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A.* **95**:14173-14178.

Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undlien DE, Cucca F, Nisticò L, Buzzetti R, Bosi E, Pociot F, Nerup J, Cambon-Thomsen A, Pugliese A, Shield JP, McKinney PA, Bain SC, Polychronakos C, Todd JA. **(1997)**. Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. *Nat Genet* **17**:350-352.

Berr A, Shafiq S, Shen WH. **(2011)** Histone modifications in transcriptional activation during plant development. *Biochim Biophys Acta* **1809**:567-576.

Bertani G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.* **62**:293-300.

Bestor TH, Verdine GL **(1994)** DNA methyltransferases. *Curr Opin Cell Biol* 6: 380 – 389.

Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, Cooke R, Lagrange T. **(2009)** RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5L elongation factor family. *EMBO Rep.* **10**:649-654.

Bird A. **(2002)** DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21.

Brosnan CA, Voinnet O. **(2011)** Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. *Curr Opin Plant Biol.* **14**:580-587.

Caikovski M, Yokthongwattana C, Habu Y, Nishimura T, Mathieu O, Paszkowski J. **(2008)** Divergent evolution of CHD3 Proteins Resulted in MOM1 Refining Epigenetic Control in Vascular Plants. *PLoS Genet* **4**:e1000165.

Calarco JP, Martienssen RA. **(2011)** Genome reprogramming and small interfering RNA in the *Arabidopsis* germline. *Curr Opin Genet Dev.* **21**:134-139.

Cao X, Springer NM, Muszynski MG, Phillips RL, Käppler S., Jacobsen SE. **(2000)** Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc Nat Acad Sci USA* **97**:4979 – 4984.

Cao X, Jacobsen SE. **(2002a)** Role of Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**: 1138 – 1144.

Cao X, Jacobsen SE **(2002b)** Locus specific control of assymetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Nat Acad Sci USA* **99**: 16491 – 16498.

Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. **(2004)** RNA silencing genes control de novo DNA methylation. *Science* **5662**:1336.

Chan SW, Henderson IR, Zhang X, Shah G, Chien JS, Jacobsen SE. **(2006)** RNAi, DRD1, and

References

Histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis. *PLoS Genet.* **2**:e83.

Chandler VL, Walbot V. **(1986)** DNA methylation of a maize transposable element correlates with loss of activity. *Proc Nat Acad Sci USA* **83**:1767-1771.

Cheng X (1995) Structure and function of DNA methyltransferases. *Annu Rev Biophys Biomol Struct.* **24**:293-318.

Chen X, Liu J, Cheng Y, Jia D. **(2002)** HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* **129**:1085-1094.

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

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM **(2012)** A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**:80-92.

Cloos PA, Christensen J, Agger K, Helin K. **(2008)** Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**:1115-1140.

Clough SJ, Bent AF **(1998)** Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**:735-743.

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

Conte S, Stevenson D, Furner I, Lloyd A. **(2009)** Multiple antibiotics resistance in *Arabidopsis* is conferred by mutations in a chloroplast-localized transporter protein. *Plant Physiol.* **151**:559-573.

Cromie KD, Ahmad K, Malik T, Buyukuslu N, Glass RE. **(1999)** Trans-dominant mutations in the 3'-terminal region of the rpoB gene define highly conserved, essential residues in the beta subunit of RNA polymerase: the GEME motif. *Genes Cells.* **4**:145-159.

Deblaere R, Bytebier B, De Greve H, Deboeck F, Schell J, Van Montagu M, Leemans J. **(1985)** Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nucleic Acids Res.* **13**:4777-4788.

Dhawan R, Luo H, Foerster AM, Abuqamar S, Du HN, Briggs SD, Mittelsten Scheid O, Mengiste T. **(2009)** HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in *Arabidopsis*. *Plant Cell* **21**:1000-1019.

Dong Q, Han F. **(2012)** Phosphorylation of histone H2A is associated with centromere function and maintenance in meiosis. *Plant J.* **71**:800-809.

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL **(2000)**. mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**:2101-2118.

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. **(2012)**

Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci USA*. **109**:E2183-2191.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, Hetzel J, Wohlschlegel JA, Pradhan S, Patel DJ, Jacobsen SE **(2012)** Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**:167-180.

Ebbs ML, Bender J **(2006)** Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell* 18:1166-1176.

El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi MA, Jacobsen SE, Cooke R, Lagrange T. **(2007)** Reiterated WG/GW motifs form functionally and evolutionary conserved ARGONAUTE-binding platforms in RNAi related components. *Genes Dev.* **21**:2539-2544.

English JJ, Jones JD. **(1998)** Epigenetic instability and transsilencing interactions associated with an SPT::Ac T-DNA locus in tobacco. *Genetics* **148**: 457–469.

Erhard KF Jr, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB. **(2009)** RNA polymerase IV functions in paramutation in *Zea mays*. *Science* **323**:1201-1205.

Eshed Y, Baum SF, Bowman JL **(1999)** Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell* **99**:199-209.

Eun C, Lorkovic ZJ, Naumann U, Long Q, Havecker ER, Simon SA, Meyers BC, Matzke AJM, Matzke M. **(2011)** AGO6 functions in RNA-Mediated Transcriptional Gene Silencing in Shoot and Root Meristems in *Arabidopsis thaliana*. *PLoS ONE* **6**:e25730.

Eun C, Lorkovic ZJ, Sasaki T, Naumann U, Matzke AJ, Matzke M. **(2012)** Use of Forward Genetic Screens to Identify Genes Required for RNA-Directed DNA Methylation in Arabidopsis thaliana. *Cold Spring Harb Symp Quant Biol*.

Finke A, Mette FM, Kuhlmann M **(2012a)** Genetic analysis of RNA-mediated gene silencing in Arabidopsis thaliana. *J. Verbr. Lebensm*. **7**:27-33.

Finke A, Kuhlmann M, Mette MF. **(2012b)** IDN2 has a role downstream of siRNA formation in RNA-directed DNA methylation. *Epigenetics*. **7**:950-960.

Finnegan EJ, Kovac KA **(2000)** Plant DNA methyltransferases. *Plant Mol Biol.* **43**:189-201.

Fischer U, Kuhlmann M, Pecinka A, Schmidt R, Mette MF. **(2008)** Local DNA features affect RNA-directed transcriptional gene silencing and DNA methylation. *Plant J*. **53**:1-10.

Fransz PF, Armstrong S, de Jong JH, Parnell LD, van Drunen C, Dean C, Zabel P, Bisseling T, Jones GH. **(2000)** Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100**:367-376.

Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, Miki D, Zhan X, Pontier D, Lagrange T, Jin H, Matzke AJM, Matzke M, Pikaard CS, Zhu JK. **(2010)** An RNA polymerase II and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**:106-110.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. **(2006)**

References

DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele specific demethylation. *Cell* **124**:495-506.

Gendler K, Paulsen T, Napoli C **(2008)** ChromDB: The Chromatin Database. *Nucleic Acid Res.* **36**:D298-302.

Goll MG, Bestor TH. **(2005)** Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**:481-514.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. **(2005)** Mechanisms of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5) methyltransferases by Dnmt3L. *J Biol Chem* **280**:13341 – 13348.

Grasser KD, Grill S, Duroux M, Launholt D, Thomsen MS, Nielsen BV, Nielsen HK, Merkle T. **(2004)** HMGB6 from *Arabidopsis thaliana* specifies a novel type of plant chromosomal HMGB protein. *Biochemistry* **43**:1309-1314.

Greenberg MV, Ausin I, Chan SW, Cokus SJ, Cuperus JT, Feng S, Law JA, Chu C, Pellegrini M, Carrington JC, Jacobsen SE. **(2011)** Identification of genes required for de novo DNA methylation in *Arabidopsis*. *Epigenetics* **6**:344-354.

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB. **(1998)** Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**:446-450.

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS. **(2012)** In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol Cell.* **48**:811-818.

Habu Y, Mathieu O, Tariq M, Probst AV, Smathajitt C, Zhu T, Paszkowski J. **(2006)** Epigenetic regulation of transcription in intermediate heterochromatin. *EMBO Rep* **7**:1279-1284.

Hagemann S, Scheer B, Schweizer D **(1993)** Repetitive sequences in the genome of Anemone blanda: identification of tandem arrays and of dispersed repeats. *Chromosoma* **102**:312-324.

Hale CJ, Stonaker JL, Gross SM, Hollick JB. **(2007)** A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize. *PLoS Biol* **5**:e275.

Haring M, Bader R, Louwers M, Schwabe A, van Driel R, Stam M. **(2010)** The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. *Plant J.* **63**:366-378.

Hartwig B, James GV, Konrad K, Schneeberger K, Turck F. (2012) Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol.* **160**:591-600.

Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC. **(2010)** The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *Plant Cell* **22**:321-344.

He XJ, Hsu YF, Pontes O, Zhu J, Lu J, Bressan RA, Pikaard C, Wang CS, Zhu JK. **(2009a)** NRPD4, a protein related to the RPB4 subunit of polymerase II, is a component of RNA polymerase IV and V and is required for RNA-directed DNA methylation. *Genes Dev.* **23**:318-330.

He XJ, Hsu YF, Zhu S, Liu HL, Pontes O, Zhu JK. **(2009b)** A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. *Genes Dev.***23**:2717-2722.

He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu HL, Wang CS, Jin H, Zhu JK.

(2009c) An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**:498-508.

Henderson IR, Deleris A,Wong W, Zhong X, Chin HG, GA Horwitz, Kelly KA, Pradhan S., Jacobsen SE. **(2010)** The De Novo Cytosine Methyltransferase DRM2 Requires Intact UBA Domains and a Catalytical Mutated Paralog DRM3 during RNA-directed DNA Methylation in Arabiopsos thaliana. *PLoS Genet.* **6**:e1001182.

Hendrich B, Bird A. **(1998)** Identification and characterization of a family of mammalian methyl CpG binding proteins. *Mol Cell Biol.* **18**:6538-6547.

Henikoff S, Comai L. **(1998)** A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. *Genetics* **149**:307-318.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. **(2005)** RNA Polymerase IV directs silencing of endogenous DNA. *Science* **308**:118-120.

Hori K, Watanabe Y. **(2005)** UPF3 suppresses aberrant spliced mRNA in Arabidopsis. *Plant J.* **43**:530-540.

Houben A, Demidov D, Caperta AD, Karimi R, Agueci F, Vlasenko L. **(2007)** Phosphorylation of histone H3 in plants--a dynamic affair. *Biochim Biophys Acta*. **1769**:308-315.

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D. **(2009)** Genome wide demethylation of *Arabidopsis* endosperm. *Science* **324**:1451-1454.

Huang L, Jones AM, Searle I, Patel K, Vogler H, Hubner NC, Baulcombe DC. **(2009)** An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol.* **16**:91-93.

Ito M, Koike A, Koizumi N, Sano H. **(2003)** Methylated DNA-binding proteins from *Arabidopsis*. *Plant Physiol*. **133**:1747-1754.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE **(2002)**. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**:556-560.

Jenuwein T, Allis CD. **(2001)** Translating the histone code. *Science* **293**:1074-1080.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. **(2007)** The SRA Methyl-cytosine-binding Domain Links DNA and Histone methylation. *Curr Biol.* **17**:379-384.

Jullien PE, Katz A, Oliva M, Ohad N, Berger F. **(2006b)** Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in Arabidopsis. *Curr Biol.* **16**:486-492.

Jullien PE, Kinoshita T, Ohad N, Berger F. **(2006a)** Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell* **18**:1360-1372.

Jullien PE, Susaki D, Yelaganula R, Higashiyama T, Berger F. **(2012)** DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Curr Biol.* **22**:1-6.

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ. **(1996)** Developmental abnormalities and epimutations associated with DNA hypermethylation mutations. *Proc Natl Acad Sci USA* **93**:12406-12411.

Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJM. **(2004)** Involvement of putative

SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol.* **14**:801-805.

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ. **(2005)** Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet.* **37**:761-765.

Kanno T, Aufsatz W, Jaligot E, Mette MF, Matzke M, Matzke AJM. **(2005)** A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep.* **6**:649-655.

Kanno T, Bucher E, Daxinger L, Huettel B, Böhmdorfer G, Gregor W, Kreil DP, Matzke M, Matzke AJM. **(2008)** A structural-maintenance-of-chromosomes hinge domain containing protein is required for RNA-directed DNA methylation. *Nat Genet.* **40**:670-675.

Kanno T, Bucher E, Daxinger L, Huettel B, Kreil DP, Matzke AJM, Matzke M. **(2009)**. RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. *EMBO Rep.* **11**:65-71.

Kareta MS, Bortello ZM, Ennis JJ, Chou C, Chedin F. **(2006)** Reconstitution and mechanism of the stimulation of de novo methylation by human Dnmt3L. *J Biol Chem.* **281**: 25893 – 25902.

Karlowski WM, Zielezinski A, Carrère J, Pontier D, Lagrange T, Cooke R. **(2010)** Genome-wide computational identification of WG/GW Argonaute-binding proteins in *Arabidopsis*. *Nucleic Acids Res.* 38:4231-4245.

Kermicle JL, Eggleston WB, Alleman M. **(1995)** Organization of paramutagenicity in R-stippled maize. *Genetics* **141**:361–372.

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL. **(1999)** Imprinting of the MEDEA polycomb gene in the *Arabidopsis* endosperm. *Plant Cell* **11**:1945-1952.

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T. **(2004)** One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**:521-523.

Kirkpatrick M. (2010) How and why chromosome inversions evolve. *PLoS Biol.* **8**:e1000501.

Köhler C, Page DR, Gagliardini V, Grossniklaus U. **(2005)** The Arabidopsis thaliana MEDEA polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**:28–30.

Kapitonov VV, Jurka J. **(2007)** Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet.* **23**:521-529.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. **(2007)** Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**:420-426.

Kornberg RD, Klug A. **(1981)** The nucleosome. *Sci Am.* **244**:52-64.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Puruganan MD, Durrant C, Mott R. **(2009)**

A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet*. **5**:e1000551.

Kozlov G, Nguyen L, Lin T, De Cresczenzo G, Park M, Gehring K. **(2007)** Structural basis of ubiquitin recognition by the ubiquitin-assoziated UBA domain of the ubiquitin ligase EDD. *J Biol Chem*. **282**:35787-35795.

Krebbers E, Hehl R, Piotrowiak R, Lönnig WE, Sommer H, Saedler H. **(1987)** Molecular analysis of paramutant plants of Antirrhinum majus and the involvement of transposable elements. *Mol Gen Genet* **209**:499-507.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG. **(1994)** The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**:1-10.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. **(2007)** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947-2948.

Law JA, Jacobsen SE. **(2010)** Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* **11**:204-220.

Law JA, Ausin I, Johnon LM, Vashisht AA, Zhu JK, Wohlschlegel JA, Jacobsen SE. **(2010)** A protein complex required for polymerase V transcripts and RNA-directed DNA Methylation in *Arabidopsis*. *Curr Biol.* **20**:951-956.

Le QH, Wright S, Yu Z, Bureau T. **(2000)** Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci USA.* **97**: 7376–7381.

Le Scouarnec S, Gribble SM. **(2012)** Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. *Heredity* **108**: 75–85.

Lermontova I, Schubert V, Fuchs J, Klatte S, Macas J, Schubert I. **(2006)** Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell* **18**:2443-2451.

Lewis JD, Meehan RR, Henzel WJ, Maurerfogy I, Jeppesen P, Klein F. **(1992)** Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**:905-914.

Li H, Durbin R. **(2009a)** Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup **(2009b)** The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**:2078-2079.

Li J, Yang Z, Yu B, Liu J, Chen X. **(2005)** Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr Biol.* **15**:1501-1507.

Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I, Jenuwein T, Khorasanizadeh S, Jacobsen SE. **(2004)** Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* **23**:4286-4296.

References

Lippman Z, May B, Yordan C, Singer T, Martienssen R. **(2003)** Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**:420-428.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. **(2009)** Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**:315-322.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. **(2008)** Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133:523-536.

Liu X, Yu CW, Duan J, Luo M, Wang K, Tian G, Cui Y, Wu K. **(2012)**. HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in *Arabidopsis*. *Plant Physiol.* **158**:119-129.

López A, Ramírez V, García-Andrade J, Flors V, Vera P. **(2011)** The RNA silencing enzyme RNA polymerase v is required for plant immunity. *PLoS Genet.* **7**:e1002434.

Lorkovic ZJ, Naumann U, Matzke AJM, Matzke M. **(2012)** Involvement of a GHKL ATPase in RNA-directed DNA methylation. *Curr Biol.* **22**:1-6.

Luo J, Hall BD. **(2007)** A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol.* **64**:101-112.

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM. **(1999)** Gene controlling fertilization independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **96**:296-301.

Luo M, Luo MZ, Buzas D, Finnegan J, Helliwell C, Dennis ES, Peacock WJ, Chaudhury A. **(2008)** UBIQUITIN-SPECIFIC PROTEASE 26 is required for seed development and the repression of *PHERES1* in *Arabidopsis*. *Genetics* **180**:229-236.

Malagnac F, Bartee L, Bender J. **(2002)** An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* 21:6842-6852.

March-Díaz R, Reyes JC. **(2009)** The beauty of being a variant: H2A.Z and the SWR1 complex in plants. *Mol Plant* **2**:565-577.

Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J. **(2007)** Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**:851-862.

Melamed-Bessudo C, Levy AA. **(2012)** Deficiency in DNA methylation increases meiotic crossover rates in eucharyotic but not in heterochromatic regions in *Arabidopsis*. *Proc Natl Acad Sci USA* **109**:E981-988.

Mentewab A, Steward CN Jr. **(2005)** Overexpression of an *Arabidopsis thaliana* ABC transporter confers kanamycin resistance to transgenic plants. *Nat Biotechnol.* **23**:177-1180.

Merkle T, Grasser KD. **(2011)** Unexpected mobility of plant chromatin-associated HMGB proteins. *Plant Signal Behav.* **6**:878-880.

Mersereau M, Pazour GJ, Das A. **(1990)** Efficient transformation of *Agrobacterium tumefaciens* by

References

---

electroporation. *Gene* **90**:149-151.

Meselson M, Yuan R. **(1968)** DNA restriction enzyme from E.coli. *Nature* **217**:1110-1114.

Mette MF, van der Winden J, Matzke MA, Matzke AJM. **(1999)** Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. *EMBO J* **18**:241-248.

Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJM. **(2000)** Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**:5194-5201.

Mette MF, Aufsatz W, Kanno T, Daxinger L, Rovina P, Matzke M, Matzke AJ. **(2005)** Analysis of double-stranded RNA and small RNAs involved in RNA-mediated transcriptional gene silencing. *Methods Mol Biol.* **309**:61-82.

Miller MJ, Barrett-Wilt GA, Hua Z, Vierstra RD. **(2010)** Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in *Arabidopsis*. *Proc Natl Acad Sci USA.* **107**:16512-16517.

Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J, Reinders J, Paszkowski J. **(2012)** Loss of DNA methylation affects the recombination landscape in *Arabidopsis. Proc Natl Acad Sci USA* **109**:5880-5885.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. **(2001)** Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis. Nature* **411**:212-214.

Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, Husmann D, Zhan Y, Lajoie BR, Patton McCord R, Hale CJ, Feng W, Micheals SD, Frand AR, Pellegrini M, Dekker J, Kim JK, Jacobsen SE. **(2012)** MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* **336**:1448-1451.

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. **(2005)** Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet.* **37**:997-1002.

Mueller TD, Feigon J. **(2002)** Solution structure of UBA domains reveal a conserved hydrophobic surface for protein-protein interaction. *J Mol Biol.* **319**: 1243-1255.

Myouga F, Tsuchimoto S, Noma K, Ohtsubo H, Ohtsubo E. **(2001)** Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet Syst.* 76:169-179.

Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D. **(1998)** Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis. Genetics* **149**:641-650.

Naumann K, Fischer A, Hofmann I, Krauss V, Phalke S, Irmler K, Hause G, Aurich AC, Dorn R, Jenuwein T, Reuter G. **(2005)** Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis. EMBO J.* **24**:1418-1429.

Nishimura T, Molinard G, Petty T, Broger L, Gabus C, Halazonetis TD, Thore S, Paszkowski J. **(2012)** Structural Basis of Transcriptonal Gene Silencing Mediated by *Arabidopsis* MOM1. *PLoS Genet.* **8**:e1002484.

Numa H, Kim JM, Matsui A, Kurihara Y, Morosawa T, Ishida J, Mochizuki Y, Kimura H, Shinozaki K,

References

Toyoda T, Seki M, Yoshikawa M, Habu Y. **(2010)** Transduction of RNA-directed DNA methylation signals to repressive histone marks in *Arabidopsis*. *EMBO J*. **29**:352-362.

Ogas J, Kaufmann S, Henderson J, Somerville C. **(1999)** PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. *Proc Natl Acad Sci USA* **96**:13839-13844.

Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Aréval E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP. **(2010)** Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* **464**:628-632.

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. **(2005)** Plant nuclear polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**:513-522.

Perruc E, Kinoshita N, Lopez-Molina L. **(2007)** The role of chromatin remodeling factor PKL in balancing osmotic stress responses during Arabidopsis seed germination. *Plant J*. **52**:927-936.

Pfaffl MW. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**:e45.

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vazquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T. **(2005)** Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Gen Dev.* **19**:2030-2040.

Pontvianne F, Blevins T, Chandrasekhara C, Feng W, Stroud H, Jacobsen SE, Michaels SD, Pikaard CS. **(2012)** Histone methyltransferases regulating rRNA gene dose and dosage control in *Arabidopsis*. *Genes Dev.* **26**:945-957.

Pósfai J, Bhagwat AS, Pósfai G, Roberts RJ. **(1989)**. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**:2421-2435.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. **(2012)** The Pfam protein families database. *Nucleic Acids Res.* **40**:D290-301.

Qin Y, Ye H, Tang N, Xiong L. **(2009)** Systematic identification of X1-homologous genes reveals a family involved in stress responses in rice. *Plant Mol Biol*. **71**:483 - 496.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. **(2000)** Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci USA* **97**:5237-5242.

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolić L, 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. *Mol Cell* **33**:192-203.

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Després B, Drevensek S, Barneche F, Dèrozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette ML, Robin S, Caboche M, Colot V. **(2011)**

References

Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* **30**:1928-1938.

Rowley MJ, Avrutsky MI, Sifuentes CJ, Pereira L, Wierzbicki AT. **(2011)** Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. *PLoS Genet.* **7**:e1002120.

Rozhon W, Baubec T, Mayerhofer J, Mittelsten Scheid O, Jonak C. **(2008)** Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. *Anal Biochem* **375**:354-360.

Salathia N, Lee HN, Sangster TA, Morneau K, Laundry CR, Schellenberg K, Behere AS, Gunderson KL, Cavalieri D, Jander G, Queitsch C. **(2007)** Indel arrays: an affordable alternative for genotyping. *Plant J.* **51**:727-737.

Scebba F, Bernacchia G, De Bastiani M, Evangelista M, Cantoni RM, Cella R, Locci MT, Pitto L. **(2003)** *Arabidopsis* MBD proteins show different binding specificities and nuclear localization. *Plant Mol Biol.* **53**:715-731.

Schubert D, Lechtenberg B, Forsbach A, Gils M, Bahadur S, Schmidt R. **(2004)** Silencing in *Arabidopsis* T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell* **16**:2561-72.

Sidorenko LV, Chandler VL. **(2008)**. RNA dependent RNA polymerase is required for enhancer mediated transcriptional silencing associated with paramutation at the maize *p1* gene. *Genetics* **180**:1983–1993.

Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL. **(2009)** A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. *PLoS Genet.* **5**:e1000725.

Singer T, Yordan C, Martienssen RA **(2001)** Robertson's mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene decrease in DNA methylation (DDM1). *Genes Dev* **15**: 591–602.

Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA. **(2009)** Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**:461-472.

Smith CL, Peterson CL. **(2005)** ATP-dependent chromatin remodeling. *Curr Top Dev Biol* **65**:115-148.

Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC. **(2007)** An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* **19**:1507-1521.

Stam M, Belele C, Dorweiler JE, Chandler VL. **(2002)** Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. *Genes Dev* **16**:1906–1918.

Steimer A, Amedeo P, Afsar K, Fransz P, Mitterlsten Scheid O, Paszkowski J. **(2000)** Endogenous

Targets of Transcriptional Gene Silencing in *Arabidopsis*. *Plant Cell* **12**:1165-1178.

Stonaker JL, Lim JP, Erhard KF Jr, Hollick JB. **(2009)** Diversity of Pol IV function is defined by mutations at the maize *rmr7* locus. *PLoS Genet* **5**:e1000706.

Strahl BD, Allis CD. **(2000)** The language of covalent histone modifications. *Nature* **403**:41-45.

Stroud H, Otero S, Desvoyes B, Ramírez-Parra E, Jacobsen SE, Gutierrez C. **(2012)** Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. **109**:5370-5375.

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. **(2013)** Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**:352-364.

Suetake I, Morimoto Y, Fuchikami T, Abe K, Tajima S. **(2006)** Stimulation effect of Dnmt3L on the DNA methylation activity of Dnmt3a2. *J Biochem* **140**: 553 – 559.

Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y **(2008)** The mechanisms selecting the guide strand from small RNA duplexes among Argonaute proteins. *Plant Cell Physiol* **49**:493-500.

To TK, Kim JM, Matsui A, Kurihara Y, Morosawa T, Ishida J, Tanaka M, Endo T, Kakutani T, Toyoda T, Kimura H, Yokoyama S, Shinozaki K, Seki M. **(2011)** Arabidopsis HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLoS Genet* **7**:e1002055.

Tricker PJ, Gibbings JG, Rodríguez López CM, Hadley P, Wilkinson MJ. **(2012)** Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. *J Exp Bot.* **63**:3799-3813.

Turner BM. (2000) Histone acetylation and an epigenetic code. *BioEssays* **22**:836-845.

Vaillant I, Schubert I, Tourmente S, Mathieu O. **(2006)** MOM1 mediates DNA methylation-independent silencing of repetitive sequences in Arabidopsis. *EMBO Rep.* **7**:1273-1278.

Vaucheret H. (2008) Plant ARGONAUTS. *Trends Plant Sci*. **13**:350-358.

Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoeppner MA, Grossniklaus U. **(1999)** Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity. *Genes Dev* **13**:2971-2982.

Vignali M, Hassan AH, Neely KE, Workman JL. **(2000)** ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol.* **20**:1899-1910.

Vongs A, Kakutani T, Martienssen RA, Richards EJ **(1993)** Arabidopsis thaliana DNA methylation mutants. *Science* **260**:1926-1928.

Walker EL, Panavas T. **(2001)** Structural features and methylation patterns associated with paramutation at the r1 locus of Zea mays. *Genetics* **159**:1201–1215.

Wassenegger M, Heimes S, Riedel L, Sänger HL **(1994)**. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**:567-576.

Wierzbicki AT, Haag JR, Pkaard CS. **(2008)** Noncoding transcription by RNA Polymerase PolIVbPolV mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**:635-648.

Wierzbicki AT, Ream T, Haag JR, Pikaard CS. **(2009)** RNA Polymerase V transcription guides

References

ARGONAUTE4 to chromatin. *Nat Genet* **41**:630-634.

Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS. **(2012)** Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev.* **26**:1825-1836.

Wöhrmann HJ, Gagliardini V, Raissig MT, Wehrle W, Arand J, Schmidt A, Tierling S, Page DR, Schöb H, Walter J, Grossniklaus U. **(2012)** Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. *Genes Dev* **26**:1837-1850.

Woo HR, Dittmer TA, Richards EC. **(2008)** Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet.* **4**:1000156.

Woo HR, Pontes O, Pikaard CS, Richards EJ. **(2007)** VIM1, a methylcytosine binding protein required for centromeric heterochromatization. *Genes Dev.* **21**:267-277.

Woodage T, Basrai MA, Baxevanis AD, Hieter P, Collins FS. **(1997)** Characterization of the CHD family of proteins. *Proc Natl Acad Sci USA* **94**:11472-11477.

Wu MF, Tian Q, Reed JW. **(2006)** Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133**:4211-4218.

Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL. **(2003)** Imprinting of the *MEA* Polycomb gene is controlled by antagonism between *MET1* methyltransferase and DME glycosylase. *Dev Cell* **5**:891-901.

Xie M, Ren G, Costa-Nunes P, Pontes O, Yu B. **(2012a)** A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Res.* **40**:4422-4431.

Xie M, Ren G, Zhang C, Yu B. **(2012b)** The DNA- and RNA-binding protein FACTOR of DNA METHYLATION 1 requires XH domain-mediated complex formation for its function in RNA-directed DNA methylation. *Plant J.* **72**:491-500.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. **(2004)** Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**:643-652

Yao Y, Bilichak A, Golubov A, Kovalchuk I. **(2012)** ddm1 plants are sensitive to methyl methane sulfonate and NaCl stresses and are deficient in DNA repair. *Plant Cell Rep.* **31**:1549-1561.

Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, Miller N, Drouaud J, Grelon M, Copenhaver GP, Mezard C, Kelly KA, Henderson IR. (2012) Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. *PLoS Genet.* **8**:e1002844.

Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, Mittelsten-Scheid O, Paszkowski J. **(2010)** MOM1 and Pol-IV/V interaction regulate the intensity and specificity of transcriptional gene silencing. *EMBO J.* **29**:340-351.

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padget RW, Steward R, Chen X **(2005)** Methylation as a

References

crucial step in plant microRNA biogenesis. *Science* **307**:932-935.

Zemach A, Grafi G. **(2003)** Characterization of *Arabidopsis thaliana* methyl CpG binding domain (MBD) proteins. *Plant J.* **34**:565-572.

Zhang CJ, Ning YQ, Zhang SW, Chen Q, Shao CR, Guo YW, Zhou JX, Li L, Chen S, He XJ. **(2012)** IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. *PLoS Genet.* **8**:e1002693.

Zhang J, Zhang F, Peterson T. **(2006)** Transposition of reversed Ac element ends generates novel chimeric genes in maize. *PLoS Genet.* **2**:e164.

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE. **(2007)** The Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci USA* **104**:4536 – 4541.

Zheng X, Zhu J, Kapoor A, Zhu JK. **(2007)** Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* **26**:1691-1791.

Zheng Z, Xing Y, He XJ, Li W, Hu Y, Yadav SK, Oh J, Zhu JK. **(2010)** An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in *Arabidopsis*. *Plant J.* **62**:92-99.

Zhu Y, Rowley MJ, Böhmdorfer G, Wierzbicki AT. **(2013)** A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Mol Cell.* **49**:298-309.

Zilberman D, Cao X, Jacobsen SE. **(2003)** ARGONAUTE4 Control of Locus specific siRNA accumulation and DNA and Histone Methylation. *Science* **299**:716-719.

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. **(2008)** Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**:125-129.

10 SUPPLEMENTARY DATA

A) *ProNOS*

```
CACACCTCTGTCGGCGGCGGGTCCCATTTGTGATAAATCCACCCATCGGATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCCCGC
CGATGACGCGGGACAAGCCGTTTTACGTTTGGAACTGACAGAACCGTAACGTTGAAGGAGCCACTGAGCCGCGGGTTTCTGGAGTTTAATGAGC
                                            Psp1406I
TAAGCACATACGTCAGAAACCATTATGCGCGTTCAAAAGTCGCCTAAGGTCACTATCAGCGTAGCAAATATTTCTTGTCAAAAATGCTCCACTGA
                                                          NheI
CGTTCCATAAATTCCCCTCGGTATCCAATTAGAGTTTGATATCACTGCTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTA
                        Alw26I
CAATTACATTTACAATTACCATGGTTGAACAAG
                      NcoI
```

B) *IGN5*

```
CACATTTGTTTAGGAAATATGTTAGTAAGACCTTGGTCCAATTGGTTCACCTGATAGTCAAAAATGTGTGGTGGTCCTTCAGGAAGAAGTCCAA
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
TCTGAAACATTTTGGGCCGTTTCTAGAAGAACTGATTGGATCAGCTGATAGAATTTTATTAAACAAAAAAAGACTATTTGTATATTGATTTCAT
TTTTAGTTCCCGAGAAGAGTAGAACAAATGCTAAAATGTATCATGCGGCCTAATAACCAACAAAACTAGTTGAAAGATGGATCAATGGTTTTTA
                                          HaeIII
GATGAAGAAAGCCCAAACCATACACTAATAATCTAATATTCTATTAAAAGGACTTGCTGTTATTCGGCCTAATAGCCAACAAAACTAATTGAAA
                                                        HaeIII
GATGGATCGGGGCTATGGAATACCTCAGAATAAAAAGTGTTATTCATTGCAGAGGACCCCTTAAGCGGACATGGTTGGGTCCTTGTTCGACAAG
ACTTAGTCCTCCATCTTGGCCTCAAGAGTGCTCGACGAAGTTTATCACCGCTTCATGCGGAATTTTATTCCTTGCTATGGGCAATAGAATGCAT
                                                          <<<<<<<<<<<<<<<<<<<<<
GAT
<<<
```

C) *IGN23*

```
TAGATTTGGTTAGGTAAAGGTTAGATTTTTATCATGCACATCTAAAAAGTCAGTACCGTGAGGACCACGACCGGTCCTGAATTATAAAAGGCCC
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
GAAAGGAAACTAGTCCCTTATTTCGTGTTTCTGAATAAAACAAATTATTAACCTTAAAACCAAATAAAATTTAGCAAACTGAAAATTGTAAACA
AAGAAACGGCACTAGATCCTACCTTGCAAAGAGGTGTCGAACGGACCGGCCTACAATGGCCATTAGTTTTAGATGGACTGCAACGGAGCAGTCCG
                                      GGCC         GGCC
CTAGGACCGCAGACAAAATAACTTTCGTCCAGGTTCGCCCTGCTAAACCCAACAAGTTTATGGACCGATCCGTAGGCTTAGTTTTTATATTTAAA
                                                          <<<<<<<<<<<<<<<<<<<<<<<
AAATAA
<<<<<<
```

D) *BASHO210*

```
GGAGATGTATATAACCTAGTAGGCCAGATCCTATTAAATATATATGTATATCGTAGAAGTATGGTTTTAGTTATCTTATTATATAAAGTATGGTTTT
>>>>>>>>>>>>>>>>>>>>>>>
CAAAGTTACTAACTCATAGGATCGTGCCACGTGTCAAGTCAAACGATCAGATGTTTACATGTGTCATTCAAAATTTAACCTTTTTTATTCTCTAA
AAAAATAGAAAATCCTTAAAAAGTAAACAAATTTACATATACTAATTTATATGTCTATATATTTATAAACAATTAACCATTTATTCAACAAAAAGA
AAATTAACCAAATCTGTGTTTAACTCATGTAATTATAAGTGAAAATAGAACCATTTTGAGAAAATAGCCTTAGCTTATAAGATTTAACTTTTCAATT
ATTTTCATTTATTAATTTTAACTCATTTAATATTAATTTGCATGATGTAAAATTAGCTTACGAGATTACGGCATGTATAGAAAATATATTTTCA
ACAATGTTTGGTTTCTAAATATTTTTATGGATTCCTGTTTTTTCTATTAATTATTAAGCCTAGGCGTTTGGGTATTCGGATCGGGTAT
                                                        <<<<<<<<<<<<<<<<<<<<<
```

Figure S1: Analyzed sequence of *TARGET-ProNOS*, *IGN5*, *IGN23* and *BASHO210*

The sequences analyzed by methylation sensitive restriction (A-C) and bisulfite sequencing (A-D) are underlined. Cytosines analyzed by bisulfite sequencing are marked black (CG), cyan (CHG) and red (CHH). Restriction sites are marked by boxes. Respective enzymes are depicted. Position of primers used for bisulfite sequencing are marked by arrows

A) *AtSN1*

GAAATTGACATGTTTTTTTTTTTTGAATATCTGGAAGTTCAGGCCCAAAGGCCTTACATGTCCCAGAGGCGGGACCCATACCAGAAATTCATTT
TCTTTGGAAAAGAGATCCTTCCCAACGTGTCTCGAATCCAGCACCTCTATCTGAGAGATTTACCACTGGGCCAACAACACGTTGGTGAAATTGA
CAT

B) *MEA-ISR*

TTTAATGTAAATATGTATTTGATGCATCTAACCATTTAGTATCTAAACAAATAAAAACAAAAAAAAGAAAAAAGCTCTTTAAAATCCGAAAGTA
ACTATTTTCAAAAAATCTAAATTATAAACTTAAATGTTTGGAATCGCGAACCGACTATTGCTAAATATAAATGCTAAATATACATGAAGATGTGA
AAAACATGTTGGATTTGTGGAATCGTTAATGACCACGGTTAAATGGCGGGATCCAAAATCCGGTTAGATTTCACAATGTCATATTACG

C) *AtMU1*

GTTTAGTGTTTATGATTATATAATTGTGTTATAATTGTTAATTATTTAGGGCTACCTATTGGAGGCCCCCTCAAATATATTGTTTCAACTGGGT
TCCACCCAAACTCTTAGCCTCTTTCTATTTCCCCCTAATCAACTTAAATTCTTCTCAATTCCACCTAATATATATAATCCCACGTAATTAACCA
ACATTAATCAAGCTGCTAAAATGCCGATTTAAAATCATAAAATCACGTTTTAAAATTAATTAGCCGTCGATTAAACCGATAATCCTACCATAATCT
ATTAACCCCAAATTGAAGCCGACTGTATGAGAAAGAGGCTTCAAATTTTCTTCTTAAAACACATCATTAGTAGCATTGTTGTTCTGTTATAAAACC
TTTGTCACGGAAAAAATTAAAATCTGAATGAAAGAGAAGGATA

D) *AtCOPIA4*

GGTTGTCTGTGCTTTCCATGGCTCAGACCTTACACTCACAACAAATTAGAAGAACGATCGAGACGGTGTGTGTTCCTCGGTTACTCTTTAACTC
AAACAGCCTACCTCTGTTTCGATGTTGAACATAAGCGACTTTACACATCTCGCCATGTCGTGTTTGATGAAGCCTCCTTTCCCTTCTCCAACCT
CACATCCCAAAATTCTCTCCCCACCGTAACCTTTGAACAGAGCTCCTCGCCGTTAGTTACGCCCATACTCTCATCATCGTCGGTTCTCCCATCT
TGTTTGTCTTCCCCGTGTACGGTCCTTCACCAACCAACCGCCGGTGACTACGCCGAAGTCACCACATTCATGACAGCCGACAACCTCACCGG
CTCCTCTGTCTCCTCACCGGTCAACCACAATGGACTTTCAAGTCCCACAGGTACGCTCTTCGTCACCCTTATTATCTTCTTCTTCATCTTTAAA
TTCTGAGCCCACTGCTCCAAATGAAAATGGGCCTGAAGCTGAGGCCCAGTCAC

Figure S2: Analyzed seqences of *AtSN1, MEA-ISR, AtMU1* and *AtCOPIA4*.

Cytosines analyzed by bisulfite sequencing are marked in black (CG), cyan (CHG) and red (CHH).

Figure S3: Protein sequence alignments of second-largest subunits of DNA dependent RNA polymerases.

Proteins of Schizosaccharomyces pombe (Sp), Drosophila melanogaster (Dm), Neurospora crassa (Nc), Mus musculus (Mm), Ceonorhaptidis elegans (Ce), Dictiostelium discoideum (Dd). ClustalW2 was used for alignments

Table S1: Primer used for detection of transgenes

Target	Primer Name	Sequence (5' ➔ 3')	Product Size (bp)
ProNOS (TARGET)	*pNOS-for-2*	ACAAGCCGTTTTACGTTTGG	427
	pNOS-rev	GGAACGTCAGTGGAGCATTT	
HPT (SILENCER)	*p19S-for-2*	AGGAACCGACAACCACTTTG	332
	Hyg-rev-2	GACATATCCACGCCCTCCT	
ProNOS-IR (1st part)	*p35S-for-2*	CGCACAATCCCACTATCCTT	677
	Spacer-rev-2	TTCACCAACTCAACCCATCA	
ProNOS-IR (2nd part)	*Spacer-for-2*	TGATGGGTTGAGTTGGTGAA	692
	Terminator-rev	CAAAGTGGTTGTCGGTTCCT	
SILENCER (Insertion)	*HinSite-for*	GAGATAGTGGAGCAATCTCTGAGATG	332
	HinSite-rev	TTCATACGAGACCCTCTGTTTTGGC	
pCMBL2 T-DNA	*MBar-Probe-for*	TGCTGTAAAGCGTTGTTTGG	673
	Mbar-Probe-rev	CTTCAGCAGGTGGGTGTAGAG	

Table S2: Sequencing Primer used in this thesis.

Primer-for	Sequence (5'➔3')	Primer-rev	Sequence (5'➔3')	Product Size (bp)
M13-for (-20)	GTAAAACGACGGCCAGT	*M13-rev (-24)*	CATGGTCATAGCTGTTTCC	
IDN2				
IDN2-for-1	AGGACGACCCAAAACAAGTG	*IDN2-rev-1*	TCACTTTCAGCTTCCCACCT	533
IDN2-for-2	CAAGATGGGAAGCACTGTGA	*IDN2-rev-2*	ATCGACCATCTTGTGCCTTT	531
IDN2-for-3	GCTTGTGTATCCTTGGAAAGGT	*IDN2-rev-3*	CACGGTAATGCTTTTGCTGA	694
IDN2-for-4	TTGCTTGTGCAGAACCTGAG	*IDN2-rev-4*	TGAGCTGTCCTTTCAACTGC	622
IDN2-for-5	GGTTGGGCTTATTTTCGTGA	*IDN2-rev-5*	CAAGAAGGCGGTCTTACCAC	694
IDN2-for-6	TCAGATTGGCATCCATTCAA	*IDN2-rev-6*	GAACCGATGCAAGACATCAA	627
NRPD2a				
NRPD2-for-1	TGCTTCGCTTAACCACTGAA	*NRPD2-rev-1*	CATTCATCAAATCCAGACAAGG	819
NRPD2-for-2	TGTGGAACCGTCTTTTGATG	*NRPD2-rev-2*	AAGAGTCATTGCACACAAGCA	808
NRPD2-for-3	TGGTGGGTATTCAAACGAAA	*NRPD2-rev-3*	TACTCATGGCCTCATGGTCA	813
NRPD2-for-4	TGGTTCATTTTATTTTCCTAATGGA	*NRPD2-rev-4*	GCACCCTTATCTCCCTCTCC	861
NRPD2-for-5	CGTATGCGGGAAAAAGAAAA	*NRPD2-rev-5*	CACCGTCTGGAGTTGACAAA	811
NRPD2-for-6	TGGTTCCTTTTCAATGTGAGC	*NRPD2-rev-6*	TGCTTCTGGGACTGGTAGAGA	857
NRPD2-for-7	ATGGGGAATCAAACAGCTTCT	*NRPD2-rev-7*	GCATGGAGAACGAACCTGTAA	806
NRPD2-for-8	AAGAATTTTGCTGCGGTTTCT	*NRPD2-rev-8*	AGATCACAAGGCTCACAGCAT	901
NRPD2-for-9	TGTATTGTGCCTGAAAAACGTC	*NRPD2-rev-9*	CGATTTCTTCGGTTCGGATA	827
DRM2				
DRM2-for-1	GTATGTGACGGTCTTTGACTCG	*DRM2-rev-1*	GTCTGAGAAGCCCATCTGAACT	846
DRM2-for-2	GCCGTGGTAAGTGATCCAAT	*DRM2-rev-2*	CAATTCAAACCTGGGATGCT	898
DRM2-for-3	GCCAGCAGCAGTAGAGGAAG	*DRM2-rev-3*	ACCAGAGGCTAAAGCCACAA	706
DRM2-for-4	GTGGCATAAGCCAAATTGC	*DRM2-rev-4*	GGCGGTTCTGGTTCTTCAT	904
DRM2-for-5	ATCCACAATCTCCCCATCAA	*DRM2-rev-5*	TCAACCCTTCGATTGTGTCA	779
DRM2-for-6	GTATTGGTGGTGGGGAAGTG	*DRM2-rev-6*	TAACCCTTCGCATGAAGGAA	627
NRPE1				
NRPE1_9-20_for	GCAGCTACTGCCATGTCAAA	*NRPE1_9-20_rev*	TCAACCAGAGACGACATCCA	825
NRPE1_11-12_for	GCTGTCAAGCAAAGTGGTGA	*NRPE1_11-12_rev*	CGTTTTCCCCAAGAACAAGA	650
AGO6				
AGO6_1-23_F	GCTTATACACCAGATAAGAGTGC	*AGO6_1-23_R*	GTCCTCCATGAGTTGGACGAAA	381
DRD1				
DRD1_2-11_for	GATGAGCTTCCTGGACTTGC	*DRD1_2-11_rev*	ATCCGCAGTACTCGTTCCAC	709

Table S3: InDel markers

Marker-ID	ass. Locus	Primer forward (5' → 3')	Primer reverse (5' → 3')	Col	L*er*	Diff
Chromosome 1						
CER 451941	*At1g13200*	AAGCCAAGTACCTCCAAGCA	GATCATCCCAAGGTCATGCT	487	408	79
CER 464650	*At1g30970*	CGACCGTCTTTAGCATTAGGAACTCTGG	TGGTTTGAATCGGTTGATTG	295	243	52
CER 470965	*At1g47310*	CCGGTGTTTTTCGTTATGGT	GGCAAGCGAGACAAAAGAGT	770	478	292
CER 461145	*At1g79420*	GGCAAAAAGGAGAGATGACG	CAATGCGCTCTGAATCTCTG	462	403	59
Chromosome 2						
Chr19	*At2g02090*	ATCGGCGTTACATACCGAAG	TCTCCCATGGTTTTCTCAGG	605	527	78
2g02770	*At2g02770*	GGCTGCCTTGTCTTTCTTTG	GGGGATCAGTCAAGTGGAAA	633	526	107
2g03090	*At2g03090*	CACGAGCATCCCTAAACAAATTTG	AAAGCATGCATGGTAATGTGAT	794	593	201
CER 460670	*At2g03500*	TATCCACGGAGGTCAAGAGG	AAATCCATATCTAAACCGGAAAAA	720	548	172
CER 466780	*At2g07690*	TTTTGAGAGACGGGGATGAC	TTTTCCCGATTCTTGTTTGC	539	493	46
CER 448739	*At2g16940*	CTCTCACGATGGTTTGAGCA	TAGGTGCACAACGTGCTCTC	671	619	52
CER 460212	*At2g37070*	GCGTAAGGACTTGGGATCAA	CTTTGTCGCCTTCGCTAATC	658	586	72
Chromosome 3						
CER 470258	*At3g09270*	GGCTTTCCCTTTGTGTGTGT	ACTGCTTACGTAGCCCCTCA	797	480	317
CER 455386	*At3g13920*	TCCCAAACTCGGATCTGAAC	TGTTTTGCTTCGTTTCAAGG	586	552	34
CER 456071	*At3g25160*	CCGGAAACCAGAAGTAACCA	TGCAAAATGCTCAGGTCAGA	583	548	35
CER 470949	*At3g29750*	CCCACAACTTCAAACGGTTC	CCACAACCAACAATGGATCA	511	390	121
CER 460928	*At3g46820*	TGCACACTCATGGTTTCCTC	TGTCTTTGTTGGGCTTACCC	513	436	77
IDMS3	*At3g49000*	GGAAAACGGGGATCTTCAAT	CATCTGGGTGTGTTCATTGG	571	545	26
CER 469892	*At3g54280*	CAGATGTCCTAACAAAGCTTGAACT	ACAAGTCCCTGCATGGCTAGAGAA	314	212	102
Chromosome 4						
CER 452235	*At4g03400*	TGCCTTTGCCTTTTAATGCT	GAAAGGTGCGTTGGGAAAT	807	592	215
CER 465765	*At4g05340*	AAACAACTAAATCGCCGTCA	CCTGTCGGTTTGCTATTGGT	469	391	78
CER 459855	*At4g11910*	AGTAGTGGGCGAGTGGAAGA	TTGCACTCATCAGGACAAGG	590	531	59
CER 452833	*At4g28085*	CTCGCAGTGGTGATGAAGAA	GCAGCTTGGTTCTGTGATGA	502	387	115
CER 453202	*At4g35700*	CCACTGCTATTTCCGTTGGT	TTCACTGTTGCGATAATGCAG	641	557	84
Chromosome 5						
5g03710	*At5g03710*	TCGAGAGACGAAACCGCTAT	TGCAGGTCTGAAGTGGAACA	676	580	96
CER 456051	*At5g05780*	GCGATGAAGGCAGCTATTGT	TTCACGAACATTACGCCATT	616	486	130
5g05930	*At5g05930*	GCAACAATAAAACCCCTTTCA	TTCCCGAGCATGAACTCTCT	788	693	95
CER 48932	*At5g06750*	CAATTGATGACGTGATTTTTGG	CAATATTGTGCCCATGCAAC	934	610	324
5g08139	*At5g08139*	TGTTCTGTTCACCGAGTCCA	CCTTCTCTGCTCCTCCTCCT	644	595	49
5g10580	*At5g10580*	TGCAAATGAAGGTGACGAAA	ACCGGTAGAGACCGTTGTTG	707	422	285
5g14180	*At5g14180*	CGACTGTCGTCTCTGTCCAC	AGATATTCCGAGTGCGATGC	782	654	128
5g15480	*At5g15480*	TCCATTTAGGCGCGAAATAC	ATGTCCTTGAGCCACTGGTC	752	659	93
5g15520	*At5g15520*	CCTTCAAACGACCAGTCTTCA	TGAAATGGCGACTGGTAAAAC	422	374	48
CER 456657	*At5g17920*	ATGCTTACTGGTCCCGTCAC	GTTGACCCTGTCTGCGATTT	662	564	98
CER 450021	*At5g25370*	TCATTGATTTCCCTCGATCA	TCTTGAAATGTGTAATTTCGGTGT	659	555	104
CER 454594	*At5g44670*	TCGGTCCTCTAACTTCTCCAA	TTTTACAAAATACAAGCCCAACAA	636	570	66
CER 454370	*At5g48905*	GCATGCCCTTGAAACTCATT	CGTGTGTGGTATGCTTTTGTTT	746	458	288
CER 457265	*At5g64560*	CACTATAGCATAGTAAGAACAGAAC	CGAAGAGGTTAGTAGTTTACT	257	173	84

Table S4: SNPs used for GoldenGate Assay.

All markers originally chosen are displayed.. Marker excluded from genotyping assays are highlighted in gray.

	Marker–ID	bp on chromosome	Polymorphism	Associated. Locus	Col/Ler allel
Chromosome 1	ILM1-1	1189377	MASC07014	AT1G04410	C/A
	ILM1-2	3502817	MASC00490	AT1G10590	G/A
	ILM1-3	6826510	MASC04085	AT1G19750	T/G
	ILM1-4	8974857	PFT1_1066	AT1G25540	C/A
	ILM1-5	11070735	MASC02650	AT1G31040	T/G
		12892644	MN1_12892627	AT1G35190	G/T
	ILM1-6	18629064	BKN000002167	AT1G50300	C/T
	ILM1-7	20728886	PERL0174752	AT1G55525	G/C
	ILM1-8	24791505	MN1_24795166	AT1G66460	C/A
	ILM1-9	28593403	MASC02069	AT1G76200	C/T
	ILM1-10	30307566	MN1_30312456	AT1G80630	C/A
Chromosome 2	ILM2-1	322693	MN2_322690	AT2G01730	C/T
	ILM2-2	2206676	PERL0290879	AT2G05801	T/A
	ILM2-3	2996531	MASC05970	AT2G07213	C/A
	ILM2-4	4256267	MASC05857	AT2G10820	T/G
	ILM2-5	7643148	MN2_7650227	AT2G17560	T/C
	ILM2-6	10933254	MASC02949	AT2G25680	T/A
	ILM2-7	13464490	MN2_13471565	AT2G31660	T/C
	ILM2-8	17114510	MASC02158	AT2G41010	G/A
Chromosome 3	ILM3-1	1689577	BKN000005423	AT3G05720	A/G
	ILM3-2	4141098	MN3_4141103	AT3G12970	C/T
	ILM3-3	7451906	MASC01672	AT3G21230	T/C
	ILM3-4	9828025	MASC01390	AT3G26740	A/G
	ILM3-5	11395730	MN3_11398208	AT3G29580	A/G
	ILM3-6	15410172	PERL0580497	AT3G43522	C/G
	ILM3-7	17370800	BKN000007115	AT3G47170	G/A
	ILM3-8	20017321	MN3_20028297	AT3G54050	C/A
	ILM3-9	21708549	MASC02322	AT3G58680	G/T
Chromosome 4	ILM4-1	270409	FRI_1888	AT4G00650	A/G
	ILM4-2	1243082	GA1_3232	AT4G02780	T/A
	ILM4-3	2502454	MASC04881	AT4G04920	G/T
	ILM4-4	5188627	MASC04672	AT4G08230	T/C
	ILM4-5	7215390	MASC00094	AT4G12040	A/T
	ILM4-6	9307139	MASC00901	AT4G16520	C/T
	ILM4-7	12244441	MASC03508	AT4G23460	G/T
	ILM4-8	14658633	MASC03154	AT4G29950	T/C
	SILENCER		SILENCER	AT4G32440	T-DNA
	ILM4-9	17442977	MN4_17442969	AT4G37000	G/T
Chromosome 5	ILM5-1	1193462	MASC04860	AT5G04280	T/C
	ILM5-2	3400853	MASC07398	AT5G10760	T/C
	ILM5-3	5388277	BKN000010233	AT5G16500	C/G
	ILM5-4	7442378	BKN000010440	AT5G22450	G/T
	ILM5-5	9987088	BKN000010668	AT5G27950	A/G
	ILM5-6	15378357	MASC04605	AT5G38410	C/T
	ILM5-7	16947516	BKN000011293	AT5G42390	A/T
	ILM5-8	19610272	MASC07356	AT5G48385	G/C
	ILM5-9	23326494	PERL1098418	AT5G57610	T/C
	ILM5-10	25043753	PERL1111442	AT5G62370	A/T

Table S5: CAPS marker

Marker-ID	associated AGI-Locus		Sequenz (5' → 3')	Size (bp)	Enzyme	Col-0 / L*er*	Restriction Fragments (bp)
C2P0305495	AT2G10602	F	TCGGCGGTTTTAGATTGATTAT	478	HpaII	T	478
		R	TGTAATCTCCTTCTGCACCTGA			C	143; 335
C2P0312504	AT2G11970	F	TTTGCGAGGGAAACAAAAAC	505	BclI	A	505;
		R	ATGCATGGGAAGAAATCGAC			T	160; 345
C2P0314702	At2g12420	F	CTCCCTAAGCTTCCCATTCC	475	ApoI	A	136; 339
		R	GCATGTATCAGGGCCAAAGT			T	93; 136; 246
C2P0319701	At2g13120	F	TGATGGAAAAACGCAATTGTA	457	EcoRI	G	185; 272
		R	TACGTGGCTAGTTCCTGGCTA			C	457
C2P0321056	At2g13300	F	CAGCGCGTTTTAAATCGTAAG	483	BsaBI	T	215; 267
		R	GCACCTTTCAAATCCTCTTCC			A	483
C2P0322823	At2g13540	F	TTGCTACCTTTTCAGTCGTCAA	483	RsaI	G	147; 336
		R	ACGCAAAAACACTAGCCCTTAG			A	483
MN36450403	At3g18730	F	CGGCGATGAAACTCATTTTT	627	HinfI	A	164; 437
		R	ACATTCGGCTTTCTCAGCAT			G	164; 171; 266
C3AB015474	At3g23300	F	AAGGGAAGTGGATTGGCTCCATG	1117	ScrFI	G	549, 569
		R	GTGTTTAGTAATGAATAATCATCA			A	1118
C3P0484614	At3g23760	F	CAAAGGGTGTGATCACCAAA	432	HinfI	T	432
		R	AACCTTTTGTGGTGGCTGAG			C	133; 299
MN38693286	At3g24070	F	CCATGTTTGGGCTAGGAAAA	600	MaeII	A	81; 189; 330
		R	ACCACATAGGCACCGACTTC			C	81; 519
C3P0609561	At3g47040	F	CCCGCACTGATATCCAAGAT	421	DdeI	A	85; 110; 226
		R	GGTGATGGTGCCTTTCAAGT			G	110; 311
C3P0611952	At3g47500	F	AACAACGGTCCACACCAACT	533	DdeI	T	229; 304
		R	AGCATCCGAGAGATGAAGGA			G	533
C3P0613960	At3g47965	F	TATGGTCCGGTTCGGTTAAA	326	BglII	T	326
		R	GCGTCAAACCCCAAAAAGTA			C	58; 264
C3P0616207	At3g48400	F	GGGAAGTGGTACGTGTGCTT	425	SacII	T	425
		R	TGAAGTTGCAAACGGAACAA			C	244; 281
C3P0617590	At3g48730	F	TCCACGGAACAAGGGAAGT	401	TaqI	G	67; 334
		R	CCTTACTGGGCCTTTCCAA			C	67; 119; 215
SGCSNP6578	At3g62097	F	TGCCTTGAAAATCTGAGAATCA	522	HinfI	A	137; 384
		R	TGCGTAGATATCGTTTGGTTTG			C	377
C5P0865146	At5g04290	F	CATCTCTGTCGAATGGCTGA	388	HpaII	T	388
		R	CTTGGAGGTTGGAGCAAGAG			C	73; 315
C5P0880455	At5g10690	F	AAGCTCCTACTGCCTTTGAGC	535	BseMI	C	535
		R	GAGGTTTGGTGGTTTGCTGTA			A	52; 483
C5-3427569	At5g10830	F	AAGCGAATCAAGAGCGAGAG	579	RsaI	A	579
		R	AAGCGAATCAAGAGCGAGAG			T	245; 334
C5-3443965	At5g10920	F	GGAGTTTGCGTCTCAAAAGG	487	VspI	G	487
		R	AAGCTTTGGGCAAATATCCA			A	204; 283
C5-4165329	At5g13120	F	GATCTTGGGCTCAAGCAAAT	644	HinfI	A	14; 630
		R	AACCCTCAAAAGCCCCTTAG			C	14; 209; 421
C5P0888834	At5g14020	F	CCAAAGCTGTACGCAACTCA	442	TaqI	T	443;
		R	CCGGATCTATCTGCTTCAGG			C	162; 281
C5P0889986	At5g14650	F	TGGTTCTACGTTCCTTGTTGG	419	FspBI	T	73; 111; 238
		R	TACTGTTGGTTTGGTCCTTGG			A	111; 311
C5-4877060	At5g15070	F	GAACGAAACGGAAAGAGGTTC	428	HinfI	C	90; 137; 227
		R	CTCCAGGTGAAAAATCACCAA			T	90; 338
C5-5154847	At5g15800	F	CACACGAGGCACAAAGGTTA	417	HincII	T	417
		R	AATTCACGCGCTTCATATCC			C	183; 234
C5-5609977	At5g17050	F	GGTTGGTACCGATCTCTTCTCTT	540	EcoRI	A	93; 447
		R	ATATTCTGACCCAACCCACAA			G	540
C5-6086382	At5g18370	F	TTGCACCGCGACATATCTAA	677	HinfI	T	152; 204; 267
		R	CAGGTTATTGCGTTGGGACT			C	267; 356

Marker-ID	associated AGI-Locus	Sequenz (5' → 3')		Size (bp)	Enzyme	Col-0 Ler	Restriction Fragments (bp)
C5-6490823	At5g19310	F	GGTGAAGATTTGCAGGAAAAA	430	HinfI	T	153; 277
		R	AACGCCATCAGTTTCCATTC			C	430
C5-6862166	At5g20320	F	CGGATGTAGACCTGGGAAGA	470	HinfI	G	148; 322
		R	ATAGGTACTCGAGGCGCTCA			A	470
C5-6899164	At5g20420	F	TCTTCTCAGCTTGCCACAGA	545	XhoI	A	545
		R	CGAGATCGGAAAACTCAAGC			T	57; 488
C5-7193937	At5g21150	F	AGGTATCGGCCGAAAGATTC	465	HhaI	T	9; 64; 392
		R	GCACTCTGGCGCAAAATAAT			C	9; 64; 107; 285
C5-7565802	At5g22750	F	AGGATTCACAAAGGCAAGACA	468	ScaI	T	167; 301
		R	TCTAACAGCTGCGTCCAATG			A	468
C5-7949680	At5g23570	F	TATGAACGCCACAAGAGCTG	491	HindIII	T	189; 302
		R	TCACCTCCAAAACCGAACTC			A	491
C5-8299411	At5g24330	F	TGATGTGGCATGATTTGGTT	402	XmnI	T	136; 266
		R	CCCTTTGCTGCTTCAATAGC			C	402
COL4	At5g24930	F	GTCCGGAGATGAACACTGGT	882	BglII	A	281; 324; 340
		R	GCATACGCTTTCCTTGAAGC			T	324; 588
C5-86467174	At5g25090	F	CGGAAAGCTCTTGAAGTTGG	444	HinfI	T	91; 353
		R	ACCACAAACCCAGCAGCTAA			C	90; 258
C5-8669116	At5g25130	F	CCAAAGGAATTCTTGCCGTA	517	Psp1406I	G	87; 133; 297
		R	CGGCAACAACTCAAAGTTCA			A	87; 430
C5-8677442	At5g25150	F	AGTTCGTCGTGGGCTACTTG	526	VspI	G	526
		R	ACGCCCATGACCTTAGTTTG			T	223; 303
C5-8697978	At5g25180	F	TGCATGAGAGGCTCTTGTCTT	488	HinfI	TA	488
		R	GCTGTGGAGAGCAGAAAAATG			AT	194; 272
C5-8824415	At5g25400	F	TTTGGTTGGGAAAACCTCTG	428	FspBI	G	428
		R	CCAAATGCCATTCCCTTATG			C	56; 372
C5-8850257	At5g25425	F	TCCGTTCGGATATTCTTTTCA	619	HinfI	C	619
		R	AAACGTTTTCATCCACTGACG			T	216; 403
C5-8902327	At5g25570	F	TAGAATCCGCCATTGTTCGT	500	XmiI	A	79; 135; 286
		R	CATCCCCATCTGAACTTGCT			G	79; 135; 233
C5-9010224	At5g25840	F	CATGGCTTGGGTAAGATGCT	425	SchI	A	14; 411
		R	GCTAGGGGTTGAACGACTCTT			G	14; 129; 282
C5-9103894	At5g26050	F	CTTCGGCTTTAGCATATTCCAC	425	RsaI	C	94; 331
		R	GGTCAAACTTTCAGAAACAGCA			T	94; 116; 331
C5-9987088	At5g27950	F	AAGCGCGGTTACAGGAAGTA	491	SpeI	A	155; 336
		R	AACCATCTTGCTCCTTGGTG			G	491

Table S6: Genotyping results of the extended mapping population of *nrd4*

The allele incidences of homozygous Col-0 (%C), Ler (%L) and the heterozygous state (%H) for all chromosome 5 specific markers tested in the total mapping population of 117 Kan^R Hyg^R M$_3$F$_2$ plants are displayed.

Marker-ID	Assoc. Locus	% C	% H	% L
PERL0865146	At5g04290	82	18	0
CER48932	At5g06750	90	9	1
5g08139	At5g08139	93	7	0
C5-3443965	At5g10920	98	2	0
C5-4165329	At5g13120	99	1	0
C5P0888834	At5g14020	24	76	0
C5-4877060	At5g15070	23	77	0
C5-5154848	At5g15800	21	79	0
C5-5609978	At5g17050	23	77	0
C5-6086383	At5g18370	98	2	0
C5-64990824	At5g19310	97	3	0
C5-6862167	At5g20320	99	1	0
C5-6899165	At5g20420	99	1	0
C5-7193938	At5g21150	98	2	0
C5-7565803	At5g22750	98	2	0
C5-7949681	At5g23570	98	2	0
C5-8299412	At5g24330	98	2	0
COL4	At5g24930	98	2	0
C5-86467175	At5g25090	98	2	0
C5-8669117	At5g25130	98	2	0
C5-8677443	At5g25150	99	1	0
C5-8697978	At5g25180	97	3	0
CER 450021	At5g25370	99	1	0
C5-8697979	At5g25400	95	5	0
C5-8850258	At5g25425	98	2	0
C5-8902328	At5g25570	98	2	0
C5-9010225	At5g25840	98	2	0
C5-9103895	At5g26050	98	2	0
C5-9987089	AT5G27950	94	6	0
BKN000010668	AT5G27950	94	6	0
MASC04605	AT5G38410	78	21	1
BKN000011293	AT5G42390	66	28	6
CER 454594	At5g44670	96	4	0
MASC07356	AT5G48385	21	63	16
PERL1098418	AT5G57610	19	55	26
PERL1111442	AT5G62370	21	50	29

n = 117

Table S7: Number of clones and sites analyzed by bisulfite sequencing in M_3 generation.

Numbers in parenthesis indicate number of sites in sequence context. Total number of analyzed sites (total), number of methylated sites (met) and percentage of analyzed sites (%) are indicated.

	K/K; H/H total	met	%	2-5 (nrd1) total	met	%	8-6 (nrd2-1) total	met	%	29-8 (nrd2-2) total	met	%	20-12 (nrd2-af) total	met	%	26-5 (nrd3-1) total	met	%	13-14 (nrd3-af) total	met	%	1-23 (nrd4-af) total	met	%	2-11 (nrd5-af) total	met	%	9-20 (nrd6-af1) total	met	%	11-12 (nrd6-af2) total	met	%	19-7 total	met	%	30-2 total	met	%
ProNOS (n)	n=18			n=20			n=15			n=21			n=12			n=17			n=19			n=14			n=12			n=21			n=11			n=19			n=20		
C (69)	1242	850	68	1380	414	30	1035	215	21	1449	443	31	828	168	20	1173	314	27	1311	255	19	966	435	45	828	108	13	1449	408	28	759	135	18	1311	875	66,7	1380	955	69,2
CG (18)	324	269	83	360	257	71	270	165	61	378	291	77	216	113	52	306	227	74	342	179	52	252	215	85,3	216	88	41	378	252	66	198	96	49	342	275	80,5	360	287	79,7
CHG (12)	216	168	78	240	102	43	180	31	17	252	103	41	144	38	26	204	66	32	228	58	25	168	88	52,4	144	15	10	252	98	39	132	30	23	228	171	75	240	187	77,9
CHH (39)	702	413	59	780	55	7	595	19	3	819	49	6	468	17	4	663	21	3	741	18	2	546	132	24,2	468	5	1	819	58	7	429	9	2	741	429	57,9	780	481	61,6
AtSN1 (n)	n=22			n=20			n=13			n=14			n=10			n=13			n=25			n=33			n=22			n=28			n=27			n=0			n=0		
C (46)	1012	408	36	966	164	17	598	78	13	644	47	7	506	31	6	322	19	6	1150	107	9	1518	186	12	1012	85	8	1288	121	9	1242	74	6						
CG (4)	88	54	60	84	54	64	52	46	89	56	23	41	44	18	41	28	8	29	100	31	31	132	70	59	88	40	46	112	61	55	108	35	32						
CHG (7)	154	95	64	147	57	39	91	25	28	98	11	11	77	8	10	49	3	6	175	11	6	231	66	27	154	22	14	196	31	16	189	14	7						
CHH (35)	770	248	32	735	53	7	455	7	2	490	13	3	385	5	1	245	8	3	875	26	3	1155	50	4	770	23	3	980	29	3	945	25	3						
MEA-ISR (n)	n=21			n=19			n=20			n=16			n=16			n=15			n=15			n=10			n=10			n=16			n=17			n.d.			n.d.		
C (35)	735	285	39	665	141	21	700	179	26	560	135	24	560	122	22	525	114	18	525	141	27	350	88	25	350	82	23	560	131	23	595	166	28						
CG (9)	189	156	83	171	122	71	180	166	92	144	128	89	144	108	75	135	107	79	135	113	84	90	77	86	90	82	91	144	114	79	153	138	90						
CHG (2)	42	9	21	38	2	5	40	0	0	32	0	0	32	1	3	30	0	0	30	0	0	20	1	6	20	0	0	32	2	6	34	2	6						
CHH (24)	504	120	24	456	17	4	480	13	3	384	7	2	384	13	3	360	7	2	360	28	8	240	9	4	240	0	0	384	15	4	408	26	6						
AtMU1 (n)	n=20			n=15			n=20			n=17			n=0			n=12			n.d.			n=11			n=16			n=0			n=0			n=0			n=0		
C (80)	1600	479	30	1200	217	18	1600	234	15	1360	261	19				960	154	16				880	149	17	1280	193	15												
CG (6)	120	95	79	90	68	76	120	94	78	102	84	82				72	53	74				66	56	85	96	66	69												
CHG (6)	120	58	48	90	42	47	120	43	36	102	52	51				72	25	35				66	26	39	96	32	33												
CHH (68)	1360	326	24	1020	107	11	1360	97	7	1156	125	11				816	76	9				748	67	9	1088	95	9												
AtCOPIA4 (n)	n=12			n=12			n=11			n=19			n=14			n=15			n=10			n=13			n=17			n=15			n=19			n=0			n=0		
C (175)	2100	567	27	2100	546	26	1925	481	25	3325	898	27	2450	603	25	2625	660	25	1750	389	22	2275	595	26	2975	737	25	2640	641	24	3325	842	25						
CG (25)	300	276	92	300	282	94	275	261	95	475	461	97	350	335	96	375	359	96	250	235	94	325	306	94	425	403	95	390	361	93	475	447	94						
CHG (22)	264	137	52	264	132	50	242	136	56	418	226	54	308	156	51	330	170	52	220	97	44	286	143	50	374	208	56	330	180	55	418	228	55						
CHH (128)	1536	138	9	1536	138	9	1408	84	6	2432	195	8	1792	112	6	1920	131	7	1280	58	5	1664	146	9	2176	126	6	1920	100	5	2432	167	7						
IGN5 (n)	n=13			n.d.			n=16			n.d.			n.d.			n.d.			n.d.			n=13			n.d.			n.d.			n.d.			n=0			n=0		
C (40)	520	214	41				640	58	9													520	185	36															
CG (3)	39	34	87				48	26	54													39	33	85															
CHG (4)	52	41	79				64	17	27													52	39	75															
CHH (33)	429	139	32				528	15	3													429	113	26															
IGN23 (n)	n=12			n.d.			n=11			n.d.			n.d.			n.d.			n.d.			n=13			n.d.			n.d.			n.d.			n=0			n=0		
C (48)	576	183	32				528	128	24													624	139	23															
CG (10)	120	77	64				110	100	91													130	108	83															
CHG (11)	132	50	38				121	22	18													143	41	29															
CHH (27)	324	56	17				297	7	2													351	6	2															
BASHO210 (n)	n=18			n.d.			n=20			n.d.			n.d.			n=15			n.d.			n=12			n.d.			n.d.			n.d.			n.d.			n.d.		
C (59)	1062	439	41				1180	176	15							885	146	17				708	133	19															
CG (9)	162	144	89				180	156	87							135	110	82				108	92	85															
CHG (4)	72	29	40				80	1	1							60	10	17				48	6	13															
CHH (46)	828	266	32				920	19	2							690	26	4				552	35	6															

Table S8: Non synonymous mutations in CDS of annotated genes in *nrd2-3*

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	*nrd2-3*			
Chromosome 1						
451743	AT1G02290	G	A	unknown protein	79	Q/*
1067411	AT1G04120	G	A	inositol hexakisphosphate	863	S/F
1127209	AT1G04230	G	A	DUF2361, unknown function	229	A/T
3388876	AT1G10330	G	A	TPR-like superfamily protein	44	V/I
3626795	AT1G10880	G	A	Putative role in response to salt stress	76	S/F
3652659	AT1G10930	G	A	DNA helicase (AtSGS1)	326	Q/*
3883900	AT1G11570	G	A	NTF2-like (NTL)	115	V/I
4156801	AT1G12240	G	A	ATBETAFRUCT4	505	A/T
5409754	AT1G15730	G	A	Cobalamin biosynthesis CobW-like protein	62	H/Y
5502831	AT1G16030	G	A	heat shock protein 70B (Hsp70b)	499	T/I
6692162	AT1G19360	G	A	arabinosyltransferase (RRA3)	17	A/V
6811581	AT1G19700	G	A	BEL family homeodomain protein (BEL10)	92	P/S
6988202	AT1G20150	G	A	Subtilisin-like serine endopeptidase protein	518	P/S
7010194	AT1G20230	G	A	PPR superfamily protein	209	E/K
7307754	AT1G20960	G	A	embryo defective 1507 (emb1507)	721	P/S
7446383	AT1G21270	G	A	serine/threonine protein kinase	413	G/R
7737465	AT1G21980	G	A	Type I PI4P 5-kinase	590	E/K
8234277	AT1G23205	G	A	invertase/pectin methylesterase inhibitor	192	A/V
8402754	AT1G23760	G	A	glycoprotein JP630	159	R/Q
9079428	AT1G26240	G	A	Proline-rich extensin-like family protein	50	P/L
9998009	AT1G28440	G	A	HAESA-like 1 (HSL1)	366	A/T
10013920	AT1G28480	G	A	glutaredoxin family (GRX480)	43	A/V
11391578	AT1G31780	G	A	COG6 domain protein	80	E/K
11913092	AT1G32870	G	A	NAC domain protein 13 (NAC13);	369	G/E
12393609	AT1G34050	G	A	Ankyrin repeat family protein	39	G/R
19479011	AT1G52310	G	A	protein kinase family protein	123	G/S
20075456	AT1G53780	G	A	peptidyl-prolyl cis-trans isomerases	326	T/I
20508200	AT1G54980	G	A	invertase/pectin methylesterase inhibitor	19	L/F
20654503	AT1G55350	G	A	maize DEK1-like	2139	S/L
20889715	AT1G55860	G	A	Uq-protein ligase	1062	L/F
21898043	AT1G59610	G	A	high molecular weight GTPase (DL3)	505	R/K
22214267	AT1G60230	G	A	Radical SAM superfamily protein	27	S/L
22623668	AT1G61330	C	T	FBD, F-box and LRR domains protein	232	L/F
22738941	AT1G61620	C	T	phosphoinositide binding	222	P/L
23349350	AT1G63010	C	T	SPX (SYG1/Pho81/XPR1) domain protein	365	V/M
23506209	AT1G63380	C	T	NAD(P)-binding Rossmann-fold protein	185	T/I
24150347	AT1G65010	C	T	Unknown protein	121	L/F
24513891	AT1G65890	C	T	acyl activating enzyme 12 (AAE12);	196	G/S
25272713	AT1G67470	C	T	Protein kinase superfamily protein	156	A/V
25669314	AT1G68460	C	T	adenylate isopentenyltransferase (IPT1)	246	G/R
26411781	AT1G70130	C	T	Concanavalin A-like lectin protein kinase	7	M/I
26586746	AT1G70520	C	T	cysteine-rich receptor-like protein kinase	197	A/T
27120847	AT1G72080	G	A	unknown protein	180	P/S
27125266	AT1G72090	G	A	Methylthiotransferase	270	A/T
27369278	AT1G72700	G	A	ATPase E1-E2 type family protein	732	G/E
28414565	AT1G75660	G	A	similar to yeast 5'-3'exonucleases (XRN3)	935	G/R
28472693	AT1G75830	G	A	plant defensin family (PDF)	61	R/Q
28884318	AT1G76900	G	A	TLP family protein	437	A/T
29064161	AT1G77330	G	A	1-aminocyclopropane-1-carboxylate oxidase	69	P/S
29350076	AT1G78060	G	A	Glycosyl hydrolase family protein	675	T/I
29547976	AT1G78560	G	A	Sodium Bile acid symporter family	172	A/V
29935557	AT1G79570	G	A	Protein kinase superfamily protein	662	Q/*
Chromosome 2						
116091	AT2G01190	C	T	Octicosapeptide/Phox/Bem1p protein	357	P/S
549280	AT2G02148	C	T	unknown protein.	13	G/R
1262223	AT2G03980	C	T	GDSL-like Lipase/Acylhydrolase protein	307	P/S
1403311	AT2G04160	C	T	AIR3	310	A/T
3051165	AT2G07360	C	T	SH3 domain-containing protein	746	A/T
5686896	AT2G13640	G	A	Transcription factor IIS family protein	132	P/L
7398676	AT2G17020	G	A	F-box/RNI-like superfamily protein	38	L/F
7710871	AT2G17750	G	A	NIP1	62	P/L
8285138	AT2G19110	G	A	HMA4	801	S/N
8849706	AT2G20560	G	A	DNAJ heat shock family protein	37	P/L
9646111	AT2G22680	G	A	Zinc finger (C3HC4-type RING finger) protein	227	E/K
9894839	AT2G23230	G	A	Terpenoid cyclase/prenyltransferase	444	G/E
10053054	AT2G23630	C	T	SKU5 similar 16 (sks16)	453	W/*

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd2-3			
15047859	AT2G35800	G	A	mitochondrial substrate carrier family protein	688	A/T
15749411	AT2G37520	G	C	Acyl-CoA N-acyltransferase	69	R/G
18824072	AT2G45690	A	T	similar to yeast Pep16p (SSE1)	284	N/K
19044121	AT2G46400	G	A	WRKY46	148	T/I
Chromosome 3						
1289653	AT3G04730	C	T	early auxin-induced (IAA16)	105	S/N
1597833	AT3G05510	C	T	Phospholipid/glycerol acyltransferase	279	S/L
2223272	AT3G07030	C	T	Alba DNA/RNA-binding protein	316	R/K
2668316	AT3G08790	C	T	ENTH/VHS/GAT family protein	56	L/F
2936378	AT3G09560	C	T	phosphatidate phosphohydrolase (PAH1)	707	D/N
2941068	AT3G09570	C	T	Lung seven transmembrane receptor	182	S/F
3344649	AT3G10690	C	T	DNA GYRASE A (GYRA)	296	E/K
3603812	AT3G11440	C	T	R2R3-MYB gene family (MYB64)	213	S/L
4047539	AT3G12730	C	T	Homeodomain-like superfamily protein	135	E/K
4556146	AT3G13840	C	T	GRAS family transcription factor	231	S/N
4584326	AT3G13898	C	T	unknown protein	106	P/S
5431231	AT3G16000	C	T	DNA-binding protein (MFP1)	664	R/K
5862311	AT3G17185	C	T	TAS3	274	T/I
5972933	AT3G17450	C	T	hAT dimerisation domain-containing protein	832	G/R
6124433	AT3G17880	C	T	thioredoxin-like disulfide reductase (TDX)	150	A/V
6734171	AT3G19420	C	T	PTEN 2 (PEN2	340	R/W
8550715	AT3G23730	C	T	XTH16	139	P/S
8570161	AT3G23780	C	T	NRPD/E2	660	E/K
9026524	AT3G24715	C	T	Protein kinase superfamily	226	Q/*
10275581	AT3G27730	C	T	DNA helicase (RCK)	804	E/K
16986452	AT3G46240	C	T	receptor protein kinase-related	338	G/E
18003005	AT3G48560	C	T	CSR1	180	A/T
23230182	AT3G62820	G	A	invertase/pectin methylesterase inhibitor	134	S/N
Chromosome 4						
95446	AT4G00230	G	A	xylem serine peptidase 1 (XSP1)	264	A/T
149400	AT4G00340	G	A	receptor-like protein kinase (RLK4)	148	G/E
515513	AT4G01220	G	A	CAZy glycosyltransferases (MGP4)	46	L/F
2679857	AT4G05200	G	A	CRK25	655	A/V
7641756	AT4G13110	G	A	BSD domain-containing protein	84	E/K
13522711	AT4G26910	A	G	Dihydrolipoamide succinyltransferase	32	S/P
17813284	AT4G37890	G	A	embryo sac development arrest 40 (EDA40	555	T/I
18260930	AT4G39210	G	A	ADP-Glucose pyrophosphorylase (APL3)	128	S/N
Chromosome 5						
111385	AT5G01270	C	T	CPL2 phosphatase	223	D/N
613832	AT5G02720	C	T	unknown protein	123	E/K
648100	AT5G02830	C	T	TPR-like superfamily protein	108	A/T
678384	AT5G02910	C	T	F-box/RNI-like superfamily protein	340	T/I
844516	AT5G03415	C	T	homolog of the animal DP protein	217	Q/*
1033467	AT5G03860	C	T	malate synthase (MLS2)	231	E/K
2930951	AT5G09420	C	T	TPR protein (TOC64-V)	497	A/V
3011137	AT5G09711	C	T	Unknown protein	4	R/K
3214601	AT5G10240	C	T	asparagine synthetase (ASN3).	268	G/D
3236961	AT5G10290	C	T	LRR transmembrane protein kinase family	212	A/T
4292264	AT5G13380	C	T	Auxin-responsive GH3 family protein	595	A/V
4576826	AT5G14200	C	T	AtIMD1	147	A/V
5433350	AT5G16590	G	A	LRR protein kinase family protein	467	G/D
6460640	AT5G19200	G	A	TSC10B	316	E/K
8264679	AT5G24290	G	A	Vacuolar iron transporter (VIT)	261	S/L
8938091	AT5G25620	G	A	YUCCA6 (YUC6)	45	A/V
9139477	AT5G26150	G	A	protein kinase family protein	138	P/S
9609599	AT5G27270	G	A	EMBRYO DEFECTIVE 976 (EMB976)	1031	A/T
15280957	AT5G38250	G	A	Protein kinase family protein	476	P/S
16221169	AT5G40480	G	A	embryo defective 3012 (EMB3012)	1401	G/E
16554801	AT5G41370	G	A	a DNA repair protein (XPB1)	568	S/N
17799694	AT5G44190	G	A	Golden2-like 2 (GLK2)	227	G/E
18177465	AT5G45050	G	A	WRKY Transcription Factor (TTR1)	1224	P/S
18628763	AT5G45930	G	A	magnesium chelatase SU (CHLI2)	152	G/E
19238276	AT5G47430	G	A	DWNN domain, a CCHC-type zinc finger	381	P/S
20406519	AT5G50150	G	A	DUF239, Unknown Function	21	V/L
24900191	AT5G61990	G	A	PPR superfamily protein	974	P/S
25208783	AT5G62760	G	A	nucleoside triphosphate hydrolase	204	P/L
25330535	AT5G63150	G	A	unknown protein	98	E/K
25378711	AT5G63320	G	A	Nuclear Protein X (NPX1)	25	P/S
25832061	AT5G64620	G	A	VIF2	63	A/T

Table S9: Non synonymous mutations in CDS of annotated genes in *nrd3-2*

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	*nrd3-2*			
Chromosome 1						
76461	*AT1G01180*	G	A	SAM-dependent methyltransferase	277	V/M
2182559	*AT1G07110*	G	A	F2KP	301	S/F
3039599	*AT1G09420*	G	A	G6PD4, similar to G-6-P dehydrogenase	396	A/V
3559411	*AT1G10710*	G	A	PHS1	172	C/Y
3725208	*AT1G11130*	G	A	LRR receptor-like kinase.	310	G/R
4032272	*AT1G11940*	G	A	beta-1,6-N-acetylglucosaminyltransferase	276	Q/*
4763574	*AT1G13940*	G	A	DUF863, unknown function	716	S/F
5224696	*AT1G15180*	G	A	MATE efflux family protein	82	C/Y
5425925	*AT1G15760*	C	T	Sterile alpha motif (SAM) domain-containing	71	P/L
5626949	*AT1G16480*	C	T	TPR-like superfamily protein	553	E/K
6191360	*AT1G17980*	C	T	poly(A) polymerase	20	G/E
8960406	*AT1G25510*	C	T	aspartyl protease	140	G/S
9131656	*AT1G26390*	C	T	FAD-binding Berberine protein	34	R/H
10266326	*AT1G29340*	C	T	E3 ubiquitin ligase	639	L/F
12400593	*AT1G34065*	C	T	S-adenosylmethionine carrier 2 (SAMC2);	70	D/N
19926347	*AT1G53400*	G	A	Ubiquitin domain-containing protein	98	E/K
20557663	*AT1G55090*	G	A	carbon-nitrogen hydrolase	604	R/K
21511512	*AT1G58090*	G	A	F-box protein	335	A/T
22039252	*AT1G59870*	G	A	ABC transporter	1273	R/K
22623449	*AT1G61330*	C	T	FBD, F-box and LRR domains containing	159	L/F
22665413	*AT1G61430*	C	T	S-locus lectin protein kinase	617	D/N
27487151	*AT1G73080*	G	A	LRR receptor kinase	880	R/K
28211223	*AT1G75160*	G	A	DUF620, unknown function	263	V/I
28366647	*AT1G75540*	G	A	B-box zinc finger transcription factor BBX21	165	G/D
28734981	*AT1G76580*	G	A	SBP domain transcription factor	96	D/N
29192966	*AT1G77680*	G	A	Ribonuclease II/R family protein;	856	S/F
29220869	*AT1G77740*	G	A	PIP5K2	80	D/N
Chromosome 2						
524732	*AT2G02090*	G	A	ETL1, CHR19	298	A/T
1138834	*AT2G03730*	G	A	ACT domain containing proteins	181	S/F
2344242	*AT2G06020*	G	A	Homeodomain-like superfamily protein;	139	R/K
7273156	*AT2G16750*	G	A	Protein kinase	314	A/T
7403065	*AT2G17033*	G	A	PPR repeat-containing protein	458	S/N
10274501	*AT2G24170*	C	T	Endomembrane protein 70 protein family;	574	A/T
11672751	*AT2G27270*	G	A	Late embryogenesis abundant (LEA) like	105	V/M
12384829	*AT2G28850*	G	A	member of CYP710A	45	P/S
13174872	*AT2G30950*	G	A	Metalloprotease	61	D/N
16677861	*AT2G39950*	G	A	unknown protein	270	L/F
Chromosome 3						
9310130	*AT3G25610*	C	T	ATPase E1-E2 type family protein	864	G/S
9851775	*AT3G26782*	C	T	TPR-like superfamily protein	358	A/V
9978327	*AT3G27050*	C	T	unknown protein	170	A/T
17458017	*AT3G47380*	C	T	pectin methylesterase inhibitor protein;	75	L/F
19979768	*AT3G53960*	C	T	Major facilitator superfamily protein	204	A/T
20712796	*AT3G55810*	C	T	Pyruvate kinase family protein	130	V/I
21394780	*AT3G57750*	C	T	Protein kinase superfamily protein	244	S/F
21652584	*AT3G58560*	C	T	DNAse I-like superfamily protein	312	E/K
21727455	*AT3G58750*	C	T	citrate synthase	2	E/K
Chromosome 4						
534181	*AT4G01270*	C	T	RING/U-box superfamily protein;	320	S/F
705397	*AT4G01650*	C	T	Polyketide cyclase/dehydrase	80	G/D
747995	*AT4G01720*	C	T	WRKY Transcription Factor (WRKY47)	416	P/S
6269002	*AT4G10020*	C	T	putative hydroxysteroid dehydrogenase	116	R/C
7016747	*AT4G11610*	C	T	lipid-binding phosphoribosyltransferase	230	A/T
7076005	*AT4G11745*	C	T	Galactose oxidase/kelch repeat protein	226	E/K
7200349	*AT4G12010*	C	T	TIR-NBS-LRR class protein	302	D/N
7686564	*AT4G13250*	C	T	chlorophyll b reductase (NYC1)	43	R/Q
8815225	*AT4G15410*	C	T	PUX5	120	P/S
12147900	*AT4G23200*	C	T	cysteine-rich receptor-like protein kinase	12	S/N
13044170	*AT4G25540*	C	T	homolog of MutS (MSH6)	735	S/N
Chromosome 5						
1622747	*AT5G05480*	C	T	Peptide-N4- asparagine amidase A	473	G/D
2530938	*AT5G07930*	C	T	a member of mei2-like gene family (MCT2)	25	P/L
2579668	*AT5G08055*	C	T	defensin-like (DEFL) family protein	38	P/S
3773722	*AT5G11710*	C	T	EPSIN1	59	S/L
3910150	*AT5G12090*	C	T	Protein kinase superfamily protein	128	P/L

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd3-2			
3978713	AT5G12300	C	T	CaLB domain family protein	307	R/K
4060550	AT5G12860	C	T	AtpOMT1	381	A/T
4293281	AT5G13390	C	T	NEF1	158	T/I
4473670	AT5G13860	C	T	ELCH-like (ELC-Like)	217	V/M
4538605	AT5G14060	C	T	lysine-sensitive aspartate kinase	476	R/C
4716112	AT5G14620	C	T	DRM2	439	W/*
4841566	AT5G14950	C	T	GMII	76	G/D
5426734	AT5G16580	C	T	beta glucosidase 2 (BGLU2)	153	D/N
6081829	AT5G18360	C	T	Disease resistance protein (TIR-NBS-LRR)	369	C/Y
6379670	AT5G19080	C	T	RING/U-box superfamily protein	257	P/S
7011635	AT5G20710	C	T	beta-galactosidase 7 (BGAL7)	245	P/L
7762114	AT5G23110	C	T	Zinc finger, C3HC4 type (RING finger)	1102	P/S
7959237	AT5G23610	C	T	Unknown protein	179	D/N
8251542	AT5G24280	C	T	GMI1	1545	E/K
11133416	AT5G29070	C	T	unknown proteins	34	G/D
15509675	AT5G38720	G	A	unknown protein	117	A/V
15743722	AT5G39320	G	A	UDP-glucose 6-dehydrogenase	157	E/K
22546269	AT5G55670	G	A	RNA-binding (RRM/RBD/RNP motifs) protein	178	P/L

Table S10: Non synonymous mutations in CDS of annotated genes in *nrd4*

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd4			
Chromosome 1						
67595	AT1G01140	C	T	CBL-interacting protein kinase, SOS2-like	START-gain	
222248	AT1G01610	C	T	Glycerol-3-phosphate acyltransferase	405	C/Y
2298806	AT1G07480	C	T	Transcription factor IIA, a/b subunit.	78	V/I
12538566	AT1G34355	C	T	Parallel Spindle 1 (PS1)	563	L/F
20478323	AT1G54930	C	T	GRF zinc finger / Zinc knuckle protein	336	V/M
22038647	AT1G59870	G	A	ATP binding cassette transporter	1101	A/T
26599707	AT1G70550	G	A	unknown function (DUF239)	406	G/D
28494145	AT1G75890	G	A	GDSL-like Lipase superfamily protein	170	E/K
28494145	AT1G75891	G	A	Potential natural antisense gene	6	S/F
29311557	AT1G77960	G	A	unknown protein	256	P/L
29995210	AT1G79710	G	A	Major facilitator superfamily protein	25	G/E
Chromosome 2						
120006	AT2G01210	C	T	LRR protein kinase family protein	552	G/R
442019	AT2G01950	C	T	LRR receptor kinase	740	E/K
1173870	AT2G03840	G	A	Member of TETRASPANIN family	198	E/K
5537513	AT2G13350	G	A	CaLB domain family protein	73	S/L
6037832	AT2G14255	G	A	Ankyrin repeat family protein	Splice Site	
8241346	AT2G19000	G	A	unknown protein	53	G/E
8387901	AT2G19385	G	A	zinc ion binding	114	A/T
10493397	AT2G24670	G	A	DUF313 protein	196	V/M
11471044	AT2G26890	G	A	GRV2	567	T/I
11817425	AT2G27710	G	A	60S acidic ribosomal protein family	70	G/R
11877561	AT2G27900	C	T	DUF2451 proteins	START gain	
11961630	AT2G28080	T	C	UDP-Glycosyltransferase	198	K/R
12489987	AT2G29080	G	A	Mitochondrial FtsH protease	785	A/V
12594811	AT2G29330	G	A	tropinone reductase (TRI)	33	G/R
13975800	AT2G32940	G	A	AGO6	230	R/*
14789435	AT2G35080	G	A	ATP binding;aminoacyl-tRNA ligases;	165	S/F
16515971	AT2G39580	G	A	Zinc finger containing protein	1253	E/K
17702800	AT2G42510	G	A	Unknown Protein	372	G/E
18501404	AT2G44860	G	A	Ribosomal protein L24e family protein	78	P/L
Chromosome 3						
380930	AT3G02130	G	A	TOADSTOOL 2/TOAD2).	69	E/K
2406692	AT3G07540	G	A	Actin-binding FH2 (formin homology 2)	258	S/L
3448756	AT3G11000	G	A	DCD (Dev. & Cell Death) domain protein	79	A/T
3787339	AT3G11960	G	A	CPSF-A subunit protein	125	G/R
18585179	AT3G50120	G	A	DUF247 protein	525	A/V
Chromosome 4						
435774	AT4G01010	C	T	Cyclic nucleotide gated channel family	382	V/I
8282709	AT4G14370	C	T	Disease resistance protein (TIR-NBS-LRR)	152	E/K
8611517	AT4G15080	C	T	DHHC-type zinc finger protein	171	E/K
9937915	AT4G17890	C	T	ARF GAP domain (AGD) protein	112	A/V

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd4			
10225605	AT4G18550	C	T	DSEL is cytosolic acylhydrolase	245	S/N
12838526	AT4G24972	C	T	TPD1	59	P/L
13522711	AT4G26910	A	G	Dihydrolipoamide succinyltransferase	32	S/P
15016031	AT4G30830	C	T	DUF593 protein	218	R/K
15709363	AT4G32551	C	T	LEUNIG	116	L/F
17567471	AT4G37360	G	A	member of CYP81D	385	P/S
17621744	AT4G37483	G	A	unknown function	26	P/L
17901719	AT4G38150	G	A	PPR superfamily protein	134	A/V
18554906	AT4G40020	G	A	Myosin heavy chain-related protein	132	S/L
Chromosome 5						
1326330	AT5G04610	G	A	SAM-dependent methyltransferases	196	A/V
2964432	AT5G09550	G	A	GDP dissociation inhibitor proteins 190	73	V/I
3819331	AT5G11850	G	A	Protein kinase	384	P/S
4145869	AT5G13060	G	A	Armadillo BTB protein	520	S/N
5410641	AT5G16560	G	A	KANADI protein (KAN)	151	S/F
6306036	AT5G18900	G	A	2-oxoglutarate (2OG) oxygenase	44	P/L
7017974	AT5G20730	G	A	auxin-regulated transcriptional activator	798	S/F
7576424	AT5G22760	G	A	PHD finger family protein	1188	G/E
7786278	AT5G23150	G	A	Putative transcription factor	36	A/T
7890038	AT5G23420	G	A	HMGB6	25	T/I
8386142	AT5G24540	G	A	beta glucosidase 31 (BGLU31)	288	A/V
8636740	AT5G25060	G	A	RRM-containing protein	428	T/I
8948426	AT5G25630	G	A	TPR-like superfamily protein	268	G/R
9436519	AT5G26820	G	A	IREG3/MAR1.	464	G/S
9626145	AT5G27310	G	A	Transcription factor IIS family protein	233	E/K
25380620	AT5G63350	G	A	unknown protein	169	R/C
26273670	AT5G65685	G	A	UDP-Glycosyltransferase	373	L/F

Table S11: Non-synonymous mutations in CDS of annotated genes in nrd5

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd5			
Chromosome 1						
1586589	AT1G05410	C	T	DUF1423 Protein	139	A/T
2018012	AT1G06590	C	T	unknown protein	689	G/R
2304855	AT1G07500	C	T	unknown protein	46	E/K
2732371	AT1G08600	C	T	ATRX	1216	A/V
4479298	AT1G13140	C	T	Xx	325	R/Q
4860639	AT1G14225	C	T	unknown protein	4	S/N
5177722	AT1G15030	C	T	CRP family protein	START gain	
5416880	AT1G15750	C	T	TPL	746	W/*
5443357	AT1G15800	C	T	unknown protein	76	W/*
5444859	AT1G15810	C	T	S15/NS1, RNA-binding protein	122	P/S
5953078	AT1G17370	C	T	oligouridylate binding protein 1B (UBP1B)	177	G/E
6386398	AT1G18560	C	T	BED zinc finger; hAT dimerisation domain	181	A/V
6701617	AT1G19390	C	T	Wall-associated kinase family protein	508	V/I
7043771	AT1G20350	C	T	mitochondrial inner membrane translocase	12	P/L
9769749	AT1G28020	G	A	TPR-like superfamily protein	220	A/V
11673877	AT1G32360	G	A	Zinc finger (CCCH-type) family protein	185	E/K
16495984	AT1G43720	G	A	unknown protein	284	R/K
17053468	AT1G45110	G	A	Tetrapyrrole (Corrin/Porphyrin) Methylases	167	A/T
17985736	AT1G48635	G	A	peroxin 3 (PEX3)	289	E/K
18462337	AT1G49870	G	A	unknown protein	451	E/K
19107015	AT1G51520	G	A	RNA-binding (RRM/RBD/RNP motifs) protein	76	G/D
19554695	AT1G52470	G	A	a/b-Hydrolases superfamily protein	105	G/R
19959370	AT1G53470	G	A	MSL4	262	E/K
24850131	AT1G66610	G	A	TRAF-like superfamily protein	229	A/V
24973050	AT1G66930	G	A	Protein kinase superfamily protein	669	C/Y
27123625	AT1G72090	T	A	Methylthiotransferase	3	D/E
27243399	AT1G72370	C	T	acidic protein associated	241	G/E
28075932	AT1G74720	G	A	Encodes a putative transmembrane protein	254	D/N
28857636	AT1G76870	G	A	unknown protein	129	M/I
29222194	AT1G77740	G	A	Encodes PIP5K2	456	G/E
29388795	AT1G78100	G	A	F-box family protein	297	D/N
29638012	AT1G78830	G	A	Curculin-like lectin family protein	166	T/I
Chromosome 2						
418495	AT2G01910	G	A	MAP65-6	209	R/K
524642	AT2G02090	G	A	ETL1/ CHR19	268	V/I

Position	Locus	Allele		Annotation	aa in protein	effect
		Col-0	nrd5			
627059	AT2G02380	G	A	GSTZ2	34	G/R
632308	AT2G02400	G	A	NAD(P)-binding Rossmann-fold protein	48	P/S
6279089	AT2G14680	G	A	maternal effect embryo arrest 13 (MEE13)	62	S/N
7100507	AT2G16390	G	A	DRD1/CHR35	693	G/D
7220905	AT2G16660	G	A	Major facilitator superfamily protein	190	A/V
12041894	AT2G28240	G	A	ATP-dependent helicase family protein	636	R/K
12262543	AT2G28610	G	A	WOX3, PRS1	143	M/I
13037863	AT2G30600	G	A	BTB/POZ domain-containing protein	152	G/R
14079822	AT2G33220	G	A	GRIM-19 protein	152	A/T
14446020	AT2G34210	C	T	Transcription elongation factor Spt5	109	P/L
16540267	AT2G39681	C	T	TAS2 encoding gene	28	D/N
17531475	AT2G42005	C	T	Transmembrane amino acid transporter	51	G/S
Chromosome 3						
175944	AT3G01460	C	T	MBD9	706	P/S
795637	AT3G03360	C	T	F-box/RNI-like protein	115	A/V
919558	AT3G03710	G	A	RIF10	6	S/N
2130148	AT3G06750	G	A	hydroxyproline-rich glycoprotein protein	21	L/F
4946401	AT3G14720	G	A	MPK19	85	V/I
5335707	AT3G15750	G	A	Yae1, N-terminal	72	A/T
19894569	AT3G53680	G	A	Acyl-CoA N-acyltransferase	543	P/S
20197323	AT3G54560	G	A	HTA11, histone H2A	90	V/M
20232513	AT3G54660	G	A	glutathione reductase	150	S/F
20266867	AT3G54750	G	A	unknown protein	283	P/S
23014671	AT3G62160	C	T	HXXXD-type acyl-transferase	317	V/M
23326591	AT3G63130	C	T	RAN GTPase activating protein	495	P/L
Chromosome 4						
12460356	AT4G23990	G	A	cellulose synthase like	705	S/N
13522711	AT4G26910	A	G	Dihydrolipoamide succinyltransferase	32	S/P
14586305	AT4G29790	G	A	unknown protein	404	G/E
15020871	AT4G30850	G	A	heptahelical transmembrane protein	249	A/V
15600744	AT4G32300	G	A	S-domain-2 5 (SD2-5)	259	G/S
16483161	AT4G34480	G	A	O-Glycosyl hydrolases family 17 protein	222	P/L
Chromosome 5						
1355547	AT5G04700	G	A	Ankyrin repeat family protein	306	Q/*
9103712	AT5G26050	C	T	Plant self-incompatibility protein S1 family;	56	D/N
15280755	AT5G38250	C	T	Protein kinase family protein	543	S/N
15995113	AT5G39960	G	A	GTP binding protein	222	G/E
16658153	AT5G41660	G	A	unknown protein	331	V/M
17246827	AT5G43000	G	A	unknown protein	126	L/F
18042261	AT5G44710	G	A	Ribosomal protein S27/S33	83	R/H
18244534	AT5G45130	G	A	RHA1	FRAME SHIFT	
19760923	AT5G48720	G	A	XRI1 (X-ray induced 1)	126	D/N
19948701	AT5G49200	G	A	ZFWD4	119	P/S
20056597	AT5G49460	G	A	ATP Citrate Lyase (ACL)	275	G/S
20428290	AT5G50175	G	A	unknown protein	50	A/T
20486735	AT5G50330	G	A	Protein kinase superfamily protein	292	R/C
21295375	AT5G52470	G	A	FIB1	188	A/T
21336826	AT5G52570	G	A	CHY	21	S/F
21867249	AT5G53860	G	A	emb2737	245	D/N
23613198	AT5G58410	G	A	HEAT/U-box domain-containing protein	751	V/I
24371208	AT5G60640	G	A	PDIL1-4	512	T/I
24637590	AT5G61260	G	A	Plant calmodulin-binding protein-related	161	R/K

Table S12: Non synonymous mutations in CDS of annotated genes in *nrd6-1*

Position	Locus	Allele		Annotation	aa in Protein	effect
		Col-0	nrd6-1			
Chromosome 1						
647756	AT1G02890	C	T	AAA-type ATPase family protein	789	V/M
1408943	AT1G04960	C	T	DUF1664 protein	215	M/I
1930794	AT1G06320	C	T	unknown protein	35	L/F
4969616	AT1G14520	C	T	MIOX1	51	D/N
4979332	AT1G14550	G	A	Peroxidase superfamily protein	74	D/N
5413624	AT1G15740	C	T	LRR family protein	390	S/F
6338863	AT1G18410	C	T	nucleoside triphosphate hydrolase protein;	654	E/K
7032404	AT1G20310	C	T	unknown protein	33	R/W

Position	Locus	Allele		Annotation	aa in Protein	effect
		Col-0	*nrd6-1*			
7408390	AT1G21160	C	T	eIF-2 family protein	1039	E/K
8129249	AT1G22960	C	T	PPR superfamily protein	332	D/N
9288943	AT1G26810	G	A	β1,3-galactosyltransferase activity protein	129	L/F
9555756	AT1G27510	G	A	DUF3506	518	T/I
10181519	AT1G29150	G	A	FUS6/COP11 interacting protein	94	A/T
16963535	AT1G44890	G	A	OXA1-like	60	A/T
17224950	AT1G46264	G	A	SCHIZORIZA	16	G/D
17538015	AT1G47670	G	A	Transmembrane amino acid transporter	216	S/L
18383497	AT1G49670	G	A	unknown protein	346	P/L
20740001	AT1G55540	G	A	unknown function	594	A/V
21482043	AT1G58050	G	A	RNA helicase family protein	951	T/I
27254628	AT1G72410	C	T	COP1-interacting protein-related	369	D/N
28770498	AT1G76660	C	T	Unknown protein	17	V/M
29771443	AT1G79130	C	T	SAUR-like auxin-responsive protein	42	A/V
30206125	AT1G80350	C	T	p60 katanin protein	374	A/T
Chromosome 2						
6754908	AT2G15470	G	A	Pectin lyase-like superfamily protein	126	G/D
7335682	AT2G16920	G	A	ubiquitin-conjugating enzyme 23 (UBC23)	839	L/F
7608167	AT2G17500	G	A	Auxin efflux carrier family protein	232	E/K
8828234	AT2G20470	G	A	AGC kinase family protein	309	H/Y
9144041	AT2G21380	G	A	Kinesin motor family protein	295	S/N
11217781	AT2G26350	G	A	PEX10	378	H/Y
14288798	AT2G33775	G	A	RALF-like protein	85	D/N
14520652	AT2G34410	G	A	Cas1p homolog protein	312	W/*
16719124	AT2G40030	G	A	NRPE1	937	G/D
Chromosome 3						
445929	AT3G02260	C	T	Calossin-like protein	623	G/R
1501797	AT3G05270	C	T	DUF869 protein	313	E/K
2147447	AT3G06810	C	T	IBR3	168	L/F
2956669	AT3G09640	C	T	APX2	90	P/S
3242875	AT3G10430	C	T	F-box domains-containing protein	330	E/K
4627100	AT3G13990	C	T	Kinase-related protein, DUF1296	500	A/T
6871451	AT3G19780	C	T	DUF179	243	L/F
7431565	AT3G21180	C	T	ACA9	1017	T/I
7681322	AT3G21800	C	T	UDP-glucosyl transferase 71B8 (UGT71B8)	122	V/M
8353181	AT3G23340	C	T	casein kinase I-like 10 (ckl10)	306	P/L
9029254	AT3G24715	G	A	Protein kinase superfamily protein	968	C/Y
9356277	AT3G25690	G	A	actin binding protein	520	G/D
17042328	AT3G46360	G	A	unknown protein	161	P/S
17622671	AT3G47770	G	A	member of ATH subfamily (ATH5)	899	G/D
23230182	AT3G62820	G	A	Plant invertase/pectin methylesterase inhibitor	134	S/N
Chromosome 4						
11151079	AT4G20820	C	T	FAD-binding Berberine family protein	307	P/L
13522711	AT4G26910	A	G	Dihydrolipoamide succinyltransferase	32	S/P
16723125	AT4G35130	C	T	TPR-like superfamily protein	125	G/E
17606469	AT4G37450	C	T	AGP18	89	S/N
18262419	AT4G39210	C	T	ADP-Glucose Pyrophosphorylase (APL3)	393	T/I
Chromosome 5						
1534887	AT5G05170	G	A	CESA3	11	P/L
1962356	AT5G06420	G	A	Zinc finger family protein	239	D/N
2571464	AT5G08010	G	A	unknown protein	518	G/S
3431227	AT5G10870	G	A	AtCM2	146	T/I
3465326	AT5G10960	G	A	Polynucleotidyl transferase	249	R/H
3869988	AT5G11980	G	A	COG complex component-related	344	L/F
4026685	AT5G12420	G	A	O-acyltransferase (WSD1-like) family protein	11	P/S
4713174	AT5G14610	G	A	DEAD box RNA helicase family protein	251	V/I
4772972	AT5G14770	G	A	TPR-like superfamily protein	909	A/V
5184394	AT5G15870	G	A	glycosyl hydrolase family 81 protein;	162	T/I
5826202	AT5G17680	G	A	disease resistance protein (TIR-NBS-LRR)	978	W/*
6500496	AT5G19310	G	A	Homeotic gene regulator (CHR23)	302	G/R
8180898	AT5G24155	G	A	FAD/NAD(P)-binding oxidoreductase protein	40	A/V
8590921	AT5G24930	G	A	CONSTANS-like 4 (COL4)	398	S/N
9598450	AT5G27240	G	A	DNAJ heat shock N-term domain cont. protein	300	G/D
15036722	AT5G37800	G	A	RHD SIX-LIKE 1 (RSL1)	176	G/S
21168501	AT5G52090	G	A	nucleoside triphosphate hydrolases	321	P/L
21946976	AT5G54080	G	A	homogentisate 1,2-dioxygenase	232	R/K
22289245	AT5G54870	G	A	unknown protein	33	V/M
23025601	AT5G56920	G	A	Cystatin/monellin superfamily protein	174	D/N

Position	Locus	Allele		Annotation	aa in Protein	effect
		Col-0	*nrd6-1*			
23800280	AT5G58940	G	A	calmodulin-binding receptor-like kinase	355	R/K
24091909	AT5G59790	G	A	DUF966 protein	215	R/K
25810610	AT5G64570	G	A	beta-d-xylosidase	658	L/F
25810610	AT5G64572	G	A	Potential natural antisense gene	105	E/K
26707270	AT5G66870	C	T	LOB domain protein	217	S/F

Table S13: Non synonymous mutations in CDS of annotated genes in *nrd6-2*

Position	Locus	Allele		Annotation	aa in Protein	effect
		Col-0	*nrd6-2*			
Chromosome 1						
463187	AT1G02330	C	T	unknown protein	80	E/K
1106873	AT1G04190	C	T	CC-TPR protein (TPR3)	299	G/R
2260027	AT1G07350	C	T	RNA-binding family protein (SR45a)	48	K/V
3193452	AT1G09830	C	T	GAR synthetase	340	D/N
3224391	AT1G09910	C	T	Rhamnogalacturonate lyase	21	A/T
3514674	AT1G10630	C	T	ARFA1F	107	L/V
3636310	AT1G10900	C	T	PI4P5-kinase family protein	8	R/K
5043501	AT1G14685	C	T	BASIC PENTACYSTEINE 2 (BPC2)	139	P/L
5332731	AT1G15520	C	T	ABC transporter	1237	E/K
5478824	AT1G15950	C	T	cinnamoyl CoA reductase	203	S/T
5499734	AT1G16020	C	T	DUF1712, unknown function	196	T/I
6042578	AT1G17580	C	T	MYA1	416	S/F
6341719	AT1G18410	C	T	nucleoside triphosphate hydrolases	124	W/*
6354100	AT1G18460	C	T	alpha/beta-Hydrolases	228	A/V
7231450	AT1G20800	C	T	F-box family protein	51	D/N
12419488	AT1G34110	C	T	LRR-like kinase	587	E/K
23390205	AT1G63080	G	A	Unknown protein	175	T/I
25735795	AT1G68560	C	T	XYL1	463	G/E
26888179	AT1G71340	C	T	PLC-like phosphodiesterases	9	R/K
26889059	AT1G71350	C	T	SUI1 family protein	452	D/N
26940863	AT1G71528	G	A	Potential natural antisense gene	697	T/M
26940863	AT1G71530	G	A	Protein kinase superfamily protein	307	R/H
Chromosome 2						
164555	AT2G01340	C	T	unknown protein	146	A/V
1450492	AT2G04235	C	T	unknown protein	360	P/L
1697449	AT2G04840	C	T	(DUF295) unknown function	127	Q/*
3035032	AT2G07310	C	T	unknown protein	186	G/E
7759259	AT2G17845	C	T	NAD(P)-binding Rossmann-fold protein	106	A/V
8007755	AT2G18470	G	A	(PERK4)	5	P/S
12851689	AT2G30105	C	T	LRR family protein	324	T/I
13829911	AT2G32590	C	T	Barren domain protein	585	V/M
14043853	AT2G33120	C	T	Synaptobrevin -like protein	118	A/T
14508222	AT2G34360	C	T	MATE efflux family protein	86	A/V
15418649	AT2G36780	C	T	UDP-Glycosyltransferase	154	G/S
16537575	AT2G39675	C	T	TAS1c gene	235	A/T
16720583	AT2G40030	C	T	NRPE1	1174	R/*
18179922	AT2G43900	C	T	Endonuclease/exonuclease/phosphatase	976	V/M
18307667	AT2G44300	C	T	lipid-transfer protein/seed storage 2S albumin	139	G/D
18710100	AT2G45403	C	T	unknown protein	138	L/F
18869312	AT2G45840	A	T	DUF821, unknown function	9	Q/H
19113064	AT2G46550	C	T	unknown protein	132	E/K
Chromosomse 3						
283032	AT3G01780	C	T	TPLATE	1055	T/I
609894	AT3G02810	C	T	Protein kinase	202	D/N
1089608	AT3G04150	C	T	RmlC-like cupins protein	191	G/E
1097994	AT3G04180	C	T	RmlC-like cupins protein	65	G/S
1727546	AT3G05800	C	T	AIF1	24	P/L
1877882	AT3G06200	C	T	nucleoside triphosphate hydrolase	69	L/F
2070479	AT3G06630	C	T	protein kinase	642	G/E
2673704	AT3G08800	C	T	SIEL	566	A/V
3287990	AT3G10530	C	T	Transducin/WD40 repeat-like protein	311	R/C
3547621	AT3G11320	C	T	Nucleotide-sugar transporter	167	A/T
3998958	AT3G12590	C	T	unknown protein	616	V/I
4053068	AT3G12750	C	T	Zrt- and Irt-related protein (ZIP)	30	S/N
6330291	AT3G18440	C	T	al-activated malate transporter (ALMT9)	479	P/L
6550057	AT3G18990	C	T	VRN1	111	D/N

Position	Locus	Allele		Annotation	aa in Protein	effect
		Col-0	*nrd6-2*			
6832613	*AT3G19670*	C	T	PRP40B	494	R/Q
8018007	*AT3G22670*	C	T	PPR superfamily protein	485	D/N
8037465	*AT3G22750*	C	T	Protein kinase	346	V/M
8125824	*AT3G22930*	*	+TTG	calmodulin-like protein (CML11)	4	I/IN
8886349	*AT3G24460*	C	T	Serinc-domain containing Ser biosynthesis	379	V/I
9088364	*AT3G24880*	C	T	Helicase/SANT-associated	1406	M/I
9811261	*AT3G26700*	C	T	Protein kinase	130	R/C
10255212	*AT3G27680*	C	T	Plant self-incompatibility protein S1 family	27	V/M
11234269	*AT3G29270*	C	T	RING/U-box superfamily protein	348	T/V
16289717	*AT3G44730*	C	T	kinesin-like protein 1 (KP1)	766	P/S
16889489	*AT3G45950*	C	T	Pre-mRNA splicing Prp18-interacting factor	377	C/Y
17218122	*AT3G46740*	C	T	Component of the TOC complex	392	R/K
17348067	*AT3G47110*	C	T	LRR protein	744	A/T
17769111	*AT3G48110*	C	T	glycine-tRNA ligase	787	A/V
19509254	*AT3G52600*	C	T	cell wall invertase 2 (CWINV2)	7	G/D
20017589	*AT3G54050*	C	T	High Cyclic Electron Flow 1 (HICEF1)	172	P/S
20176658	*AT3G54500*	C	T	unknown protein	467	G/E
21342854	*AT3G57630*	C	T	exostosin family protein	76	R/H
23230182	*AT3G62820*	G	A	Invertase/pectin methylesterase inhibitor	134	S/N
Chromsome 4						
286996	*AT4G00700*	G	A	lipid-binding phosphoribosyltransferase	569	G/E
843243	*AT4G01940*	G	A	NFU1	49	T/M
1218879	*AT4G02740*	G	A	F-box/RNI-like superfamily protein;	294	R/C
2534678	*AT4G04960*	C	T	Concanavalin A-like lectin protein kinase	528	A/V
6015932	*AT4G09490*	C	T	Polynucleotidyl transferase	55	C/Y
6310305	*AT4G10110*	C	T	RNA-binding (RRM/RBD/RNP motifs) protein	113	A/T
13522711	*AT4G26910*	A	G	Dihydrolipoamide succinyltransferase;	32	S/P
16430890	*AT4G34350*	G	A	ISPH	50	S/L
17415931	*AT4G36930*	G	A	SPT, bHLH transcription factor	370	D/N
18025954	*AT4G38552*	G	A	Potential natural antisense gene	813	P/S
18279192	*AT4G39270*	G	A	LRR protein kinase	693	D/N
Chromosome 5						
273274	*AT5G01730*	C	T	Potential natural antisense gene	1127	G/E
273274	*AT5G01732*	C	T	SCAR4	146	P/S
920099	*AT5G03620*	C	T	Subtilisin-like serine endopeptidase	284	S/L
2133030	*AT5G06865*	C	T	Potential natural antisense gene	190	R/K
2133030	*AT5G06860*	C	T	polygalacturonase inhibiting protein (PGIP1)	197	L/F
2165350	*AT5G06970*	C	T	Munc13 homology 1	85	S/N
2335270	*AT5G07380*	C	T	unknown protein	487	A/V
2641533	*AT5G08210*	C	T	microRNA of unknown function	48	V/M
3450753	*AT5G10940*	C	T	transducin / WD-40 repeat protein (ASG2)	488	A/T
5811233	*AT5G17640*	C	T	DUF1005, unknown function	371	M/I
6029762	*AT5G18245*	C	T	Potential natural antisense gene	305	L/F
6149007	*AT5G18525*	C	T	protein serine/threonine kinases	1163	W/*
10785538	*AT5G28750*	C	T	Bacterial mttA/Hcf106	47	R/H
14763987	*AT5G37290*	C	T	ARM repeat superfamily protein	144	P/S
15108753	*AT5G37940*	C	T	Zinc-binding dehydrogenase	332	S/F
16802668	*AT5G42010*	C	T	Transducin/WD40 repeat-like protein;	130	A/V
16809618	*AT5G42020*	C	T	Luminal binding protein (BiP2)	164	E/K

PUBLICATIONS CONNECTED WITH THE SUBMITTED THESIS

Finke A, Mette FM, Kuhlmann M **(2012)** Genetic analysis of RNA-mediated gene silencing in *Arabidopsis thaliana*. *J. Verbr. Lebensm*. **7**:27-33.

Finke A, Kuhlmann M, Mette MF. **(2012)** IDN2 has a role downstream of siRNA formation in RNA-directed DNA methylation. *Epigenetics* **7**:950-960.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt, dass diese Dissertation von mir bisher weder der Naturwissenschaftlichen Fakultät I der Martin-Luther-Universität Halle-Wittenberg noch einer anderen wissenschaftlichen Einrichtung zum Zweck der Promotion eingereicht wurde.

Ich erkläre ferner, dass ich diese Arbeit selbständig und nur unter Zuhilfenahme der angegebenen Hilfsmittel und Literatur angefertigt und wörtlich oder inhaltlich entnommene Stellen sind als solche kenntlich gemacht habe.

Datum: Unterschrift:

                                                                      Andreas Finke

CURRICULUM VITAE

Personal data

Name	Andreas Finke
Date of Birth	09/11/1981
Place of Birth	Wolfen
Nationality	German
Martial Status	unmarried
Contact information	Barfüßerstraße 11
	06108 Halle (Saale)
	phone: +49(0)174 42 52 248
	e-mail: Andreas.Finke.MLU@web.de

School education

1988 - 1992	Grundschule Sandersdorf
1992 - 2000	"Anne Frank" Gymnasium Sandersdorf

University studies

2001 - 2003	Studies of "Information technology" at the Technical University Dresden, Germany
2003 - 2008	Studies of biochemistry at the Martin–Luther–University Halle Wittenberg, Germany

Diploma thesis: "Identification and characterization of target of the epigenetic regulator TFL2 in *A. thaliana*" (Prof. C. Wasternack, Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany)

Positions held

2006 – 2007	Research assistant of Prof. Dr. C. Wasternack, Leibniz Institute of Plant Biochemistry, Martin – Luther - University Halle-Wittenberg, Germany
2009 – 2012	PhD student in Research Group Epigenetics (Supervisor: Dr. Michael Florian Mette), Department of Cytogenetics and Genome Analysis of Prof. I. Schubert, IPK Gatersleben, Germany

PhD thesis: "The genetic Characterization of RNA-directed Transcriptional Gene Silencing in *Arabidopsis thaliana*"

Halle (Saale), April 2013

Andreas Finke