Expression of *UGT76E12* and *UGT76E11*, two UDP-glycosyltransferases involved in biotic and abiotic stress responses, is differentially regulated by their antisense lncRNAs

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlich Fakultät I – Biowissenschaften – der

Martin-Luther-Universität

Halle-Wittenberg,

vorgelegt

von Herr

Ammar Mousa Jaber

geb. Am 10.05.1988 in San Francisco

Gutachter:

Prof. Dr. Stefan Hüttelmaier

Prof. Dr. Sven-Erik Behrens

Prof. Dr. Dorothee Staiger

Day of defense: Halle (Saale), 15.07.2020

Table of Contents

1 Abstract 1
2 Introduction 2
2.1 RNA 2
2.1.1 Functions and Features 2
2.1.2 Long non-coding RNAs 3
2.1.3 LncNATs in plants 4
2.1.4 Epigenetic Gene Regulation 5
2.2 Uridine diphosphate glycosyltransferases (UGTs) 10
2.3 Abiotic and biotic stress 13
2.4 B. cinerea 14
3 Thesis Objectives 17
4 Results 18
4.1 Three independent natural antisense transcripts overlap the UT76E11 and UGT76E12 genes 18
4.2 NAT-UGT76E12 expression is higher than UGT76E12 at different developmental stages 20
4.3 NAT-UGT76E12 and UGT76E12 are expressed in roots. 21
4.4 Increased NAT-UGT76E12 expression in cis but not trans affects UGT76E12 expression 23
4.5 Knockout of NAT-UGT76E12 results in decreased UGT76E12 expression 26
4.6 UGT76E12 RNA stability is unaffected in the absence of NAT-UGT76E12 28
4.7 Absence of NAT-UGT76E12 results in altered UGT76E12 DNA methylation 29
4.8 NAT-UGT76E12 knockout has reduced amount of histone modifications at the 3’ end of the UGT76E12 locus 30
4.9 UGT76E12 induction is independent of NAT-UGT76E12 expression 33
4.10 Salt stress increases the expression of the sense and antisense genes 34
4.11 Alterations of UGT76E11 and UGT76E12 expression affects susceptibility to B. cinerea infection 35
5 Discussion 39
5.1 Transcript characterization 39
5.2 Promoter activity and expression 41
5.3 Function and potential mechanism of NAT-UGT76E12 action 43
5.4 UGT76E subfamily 48
5.5 Functions of UGT76E12 and UGT76E11 48
6 Summary and conclusions 51
7 Materials and Methods 53
7.1 Gateway cloning and plasmid vectors 53
7.2 Plant lines 53
7.3 In vitro growth conditions 53
7.4 Liquid culture growth conditions 54
7.5 A. thaliana transformation 54
7.6 Selection of transgenic plants 54
7.7 RNA isolation 55

7.8 Complementary DNA synthesis (cDNA) 55
7.9 Polymerase Chain Reaction (PCR) 56
7.10 Quantitative Real Time PCR (qRT-PCR) 56
7.10.1 Data analysis 57
7.11 Rapid amplification of cDNA ends (RACE) 58
7.11.1 3'RACE 58
7.11.2 Circular 5'RACE 58
7.12 Histochemical staining 60
7.13 Overexpression constructs 61
7.14 CRISPR/Cas9 Editing 62
7.15 Proteomics 64
7.15.1 Protein extraction 64
7.15.2 SDS-PAGE 64
7.15.3 Coomassie staining 65
7.15.4 Protein detection 65
7.16 RNA stability assay 66
7.17 DNA methylation assay 67
7.17.1 Growth conditions 67
7.17.2 DNA methylation analysis 67
7.18 Chromatin Immunoprecipitation (ChIP) 68
7.18.1 Growth conditions 68
7.18.2 Crosslinking 68
7.18.3 Sonication 68
7.18.4 Immunoprecipitation 68
7.18.5 Quantification and data analysis 69
7.19 T-DNA insertion lines 70
7.20 Time course experiments 70
7.20.1 Growth conditions 70
7.20.2 Collection and analysis 71
7.21 Salt stress 71
7.21.1 Growth Conditions 71
7.21.2 Treatment 71
7.22 B. cinerea infection assay 72
7.22.1 A. thaliana growth 72
7.22.2 B. cinerea growth 72
7.22.3 Spore preparation 72
7.22.4 Infection 72
7.22.5 Lesion area and DNA quantification 73
7.22.6 Graphing and statistics 74
7.23 Statistics and graphs 75
8 Supplementary Figures 76
9 References 89

1 Abstract

In recent years, eukaryotic non-coding RNAs were found to play an important role in a variety of processes ranging from gene expression regulation to translation. In plants, they have been shown to be involved in a wide range of biological processes including flowering time regulation, root development, and hormone and stress responses. Natural antisense long non-coding RNAs (lncNATs), a sub-type of lncRNAs that are transcribed from the complementary DNA strand of a protein-coding gene, have been revealed to influence development in various plant species. Using the model organism *Arabidopsis thaliana*, we investigated the regulatory potential of a lncNAT which overlaps the *UDP-GLYCOSYLTRANSFERASE 76E12* gene (*UGT76E12*). Basal expression of *UGT76E12* and its lncNAT (*NAT-UGT76E12*) is developmentally regulated and mostly localized in the root. Additionally, the expression of *UGT76E12* is strongly induced in response to salt stress and to necrotrophic fungal infection. Ectopic overexpression of *NAT-UGT76E12* did not affect sense gene expression, however, two independent knockout lines of *NAT-UGT76E12* (*nat-ugt76e12*) showed reduced steady-state levels of *UGT76E12* mRNA. In addition, the *nat-ugt76e12* knockout had decreased levels of *UGT76E12* transcripts in response to stress. The absence of differences in *UGT76E12* mRNA stability between *WT* and *nat-ugt76e12* lines and the analogous localization of promoter activity in reporter gene lines of *UGT76E12* and *NAT-UGT76E12*, suggest a *cis* regulatory mechanism mediated by the NAT. Chromatin immune precipitation (ChIP) and DNA methylation analysis indicate an altered chromatin state in *nat-ugt76e12* compared to the *WT* confirming the predicted *cis* regulatory role of *NAT-UGT76E12* on *UGT76E12* expression. Furthermore, the expression levels of *UGT76E11*, the closest related homolog located just upstream of *UGT76E12*, are not altered when *NAT-UGT76E12* is overexpressed or knocked out indicating that this regulatory effect is gene-specific. Additionally, changes in expression levels of *NAT-UGT76E11*, a lncNAT overlapping *UGT76E11*, do not influence the expression of its sense gene.

We also investigated the role of *UGT76E12* and *UGT76E11* in biotic and abiotic stress. Overexpression and knockout resulted in increased and decreased susceptibility to infection with *Botrytis cinerea*, respectively. In contrast, additional experiments showed that overexpression conferred enhanced tolerance to salt stress while the knockout resulted in a decrease in tolerance.

2 Introduction

2.1 *RNA*

2.1.1 Functions and Features

The central dogma of molecular biology proposed that messenger RNA (mRNA) is transcribed from DNA which carries genetic information that directs the synthesis of proteins (Voet, Voet, & Pratt, 2016). While mRNAs have a critical function in eukaryotes, most RNAs do not code for proteins and are able to function as structured RNAs (Doolittle, 2013). These RNAs are called noncoding (ncRNAs) and have been shown to have a multitude of functions, including transcription regulation, RNA processing, and enzymatic activity (**Figure 2.1**). Noncoding RNAs encompass a diverse set of transcripts, including housekeeping ncRNAs (ribosomal RNAs, transfer RNAs, small nuclear RNAs, and small nucleolar RNAs), small ncRNAs (miRNAs and siRNAs), and long ncRNAs (ncRNAs larger than 200 nucleotides in length). RNA binding proteins (RBP) can interact with ncRNAs and regulate RNA processing, modification, stability, translation, and localization. They can also act on RBPs influencing the protein's function, interactions, stability, and localization (Hentze, Castello, Schwarzl, & Preiss, 2018).

Since RNA contains ribose, characterized by the presence of the 2'-hydroxyl group on the pentose ring, it is more susceptible to hydrolysis than DNA (deoxyribonucleic acid) and results in reduced stability. In addition, RNA contains the unmethylated form of the base thymine, called uracil (U). Unlike DNA, RNA is mostly found as a single strand and can assume highly diverse secondary structure. The presence of internal self-complementary sequences can form hairpin loops, bulges, and internal loops greatly affecting its structure and functionality (Voet et al., 2016).

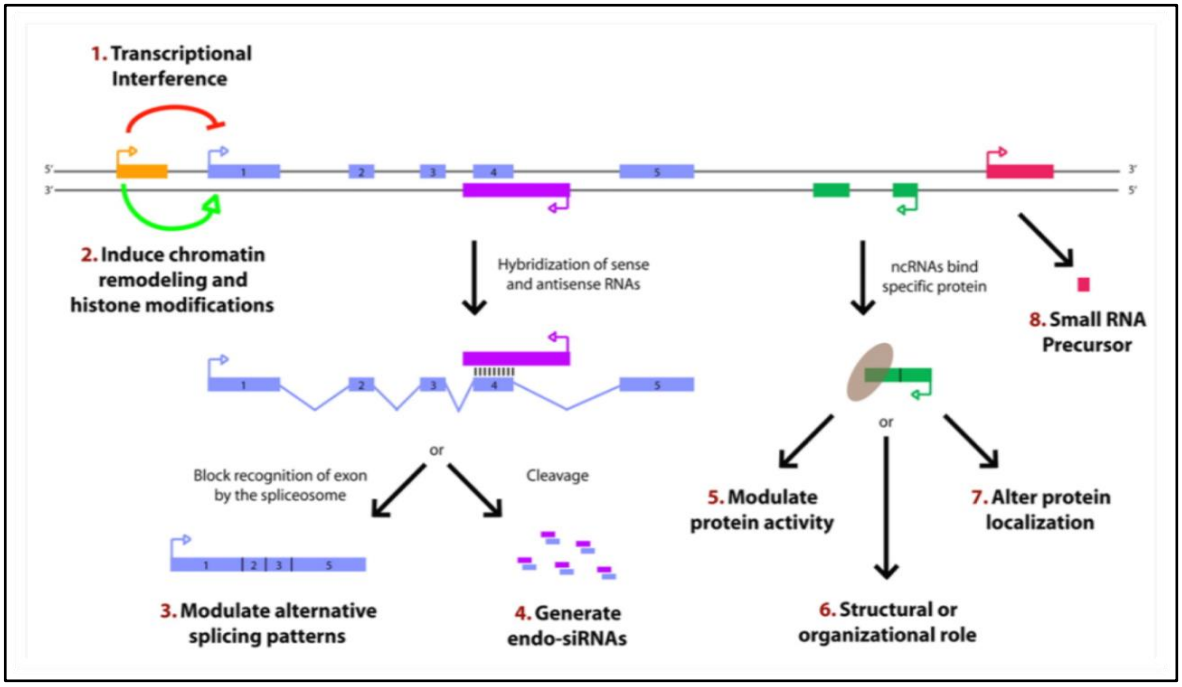


Figure 2.1 Mechanism of action of long non-coding RNAs. ***1,*** Transcription of a lncRNA can up- or down-regulate adjacent genes. ***2,*** Recruitment of histone modifying complexes that affect chromatin methylation and acetylation status. ***3,*** Sense and antisense transcripts can hybridize and modulated exon splicing. ***4,*** Hybridized RNAs are also recognized by Dicer-like proteins (DCLs) and generate endogenous small interfering RNAs (endo-siRNAs) which can subsequently be associated with RNA-induced silencing complexes (RISC) to mediate RNA interference (RNAi). ***5,6,7,*** By binding proteins lncRNAs Influence protein activity, structure, localization. ***8,*** lncRNAs are precursors of small RNAs. Legend and figure adapted from (Wilusz, Sunwoo, & Spector, 2009)

2.1.2 Long non-coding RNAs

In recent years, long non-coding RNAs (lncRNA) have been shown to have a role in a wide variety of molecular mechanisms which include: interacting with proteins and miRNAs, influencing mRNA stability and translation, and modulating chromatin modifications (Ariel, Romero-Barrios, Jegu, Benhamed, & Crespi, 2015). LncRNAs can be classified into four groups based on the basis of their genomic origins: 1) long intergenic ncRNAs (lincRNAs), 2) intronic ncRNAs (incRNAs), 3) promoter lncRNAs, and the topic of this thesis work, 4) natural antisense transcripts (lncNATs) transcribed from the complementary DNA strand of the associated genes (**Figure 2.2**). LncNATs can be further classified into sub-types depending on the degree and location of complementarity: 5' overlapping, 3' overlapping, complete overlapping, or other (partial overlap) (Mattick & Rinn, 2015; Yin et al., 2007).

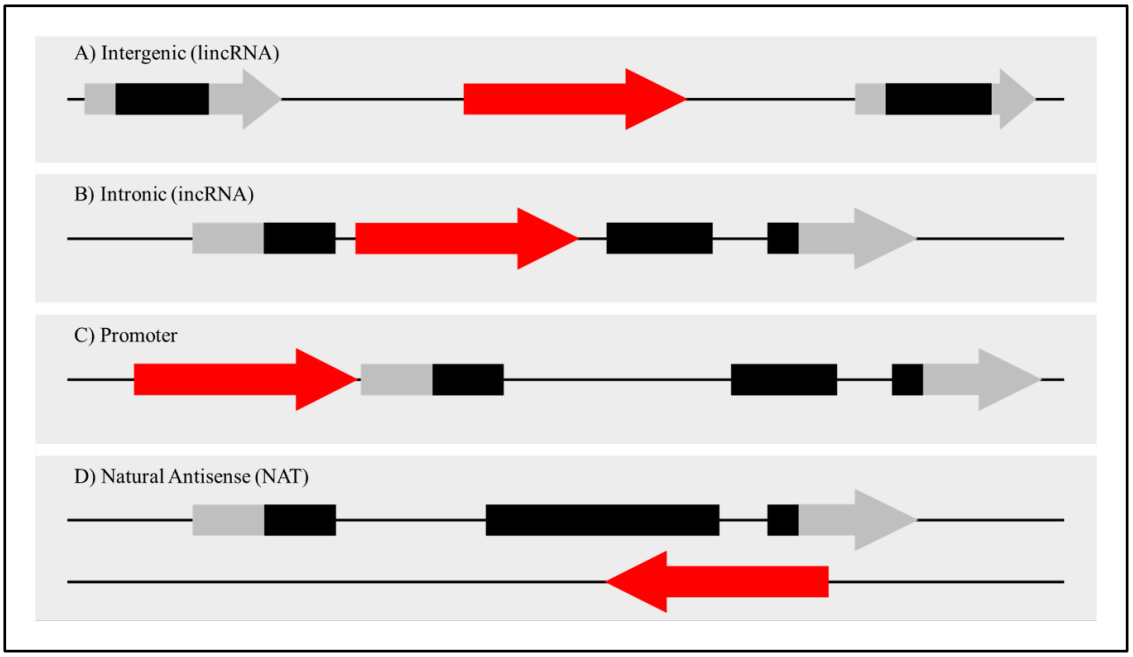


Figure 2.2 Types of long non-coding RNAs. ***A,*** Intergenic, at least 1kb away from nearby genes. ***B,*** Intronic, transcribed from an intronic region of another gene. ***C,*** Promoter, located in the promoter region (~1kb) upstream of the 5' end. ***D,*** Natural Antisense, overlapping and complementary to a sense gene. Except for NATs, lncRNAs can be in the sense or antisense orientation. Red arrows indicate the lncRNAs. Exons and untranslated regions (UTR) of coding genes are represented in black and grey, respectively. Figure based on *Ariel et al*., 2015.

2.1.3 LncNATs in plants

Studies in eukaryotes have revealed that over 90% of the genome is transcribed generating mostly non-coding RNAs. Transcriptome data of *A. thaliana* identified over 30,000 lncNATs (between 200–12,370 nt) and 70% of protein coding genes are potentially associated with a lncNAT (H. Wang et al., 2014). Approximately 60% NAT pairs show complete overlap while others are only complementary at 5′ or the 3′ ends.(H. Wang et al., 2014). In plants, relatively few lncNATs have been characterized and the mechanisms of action of many are still not fully understood. However, they have been shown to be involved in a wide range of biological processes including: flowering time regulations, root development, and hormone and stress responses (Heo, Lee, & Sung, 2013; H. V. Wang & Chekanova, 2017; J. Wang, Meng, Dobrovolskaya, Orlov, & Chen, 2017).

2.1.3.1 Transcription

Similar to mammals, the transcription of a majority of lncRNAs is mediated by RNA polymerase II (RNA Pol II) and as a result they possess a 5' cap structure and are 3' polyadenylated suggesting they are potentially exported into the cytoplasm (Chekanova, 2015; Y. Huang & Carmichael, 1996). Some are transcribed by two plant exclusive RNA

polymerases, RNA Pol IV and V, and lack a 5' cap and polyadenylation, therefore, remain and function within the nucleus (Wierzbicki, Haag, & Pikaard, 2008).

2.1.3.2 Gene expression regulation

Regulation of gene expression by lncNATs can occur at the site of its transcription, influencing the complementary gene (*cis*), or at another gene locus (*trans*) (H. V. Wang & Chekanova, 2017). Of the studied plant lncNATs, all act on their sense gene and are involved in important aspects of expression modulation. Two lncNATs have been shown to be involved in a negative feedback loop with their sense gene: gametophyte development and vegetative growth have been shown to be influenced by *Antisense Heat Stress Factors B2a* (Wunderlich, Gross-Hardt, & Schoffl, 2014) while another, *FLORE*, influences circadian oscillations and subsequently the onset of flowering (Henriques et al., 2017). In *Oryza sativa*, *antisense PHO1;2* when upregulated due to phosphate starvation results in higher PHO1;2 protein levels by increasing translation of *PHO1;2* mRNA (Jabnoune et al., 2013). The *DELAY OF GERMINATION* gene (*DOG1*), involved in the inhibition of seed germination, is downregulated by a*ntisense DOG1* allowing seeds to germinate (Fedak et al., 2016). And finally the most studied and well understood lncNAT *COOLAIR*, which after a period of vernalization, accumulates resulting in the silencing of *FLOWERING LOCUS C* (*FLC*) transcription controlling appropriate flowering time (Csorba, Questa, Sun, & Dean, 2014; Swiezewski, Liu, Magusin, & Dean, 2009). However, from the above-mentioned examples only the mechanism of *COOLAIR* was discovered. The Dean group showed *COOLAIR* has a conserved structure that can recruit POLYCOMB REPRESSIVE COMPLEX (PRC2) to the locus of *FLC* and mediate trimethylation of lysine 27 on histone 3, a repressive modification, resulting in silencing of *FLC* expression allowing the plant to flower (Csorba et al., 2014; Hawkes et al., 2016). Using fluorescence *in situ* hybridization (FISH), they further concluded that while both can be expressed in the same cell, they are not expressed simultaneously at the same locus, indicating a *cis* regulatory mechanism. Active transcription of *COOLAIR* can repress the transcription of *FLC* and while one cell can express both, they are mutually exclusive at individual loci and are always transcribed from different alleles (Rosa, Duncan, & Dean, 2016).

2.1.4 Epigenetic Gene Regulation

Epigenetics comprises a variety of mechanisms that regulate gene expression and chromatin structure. They include heritable and non-inheritable modifications to DNA and

histones that can change depending on environmental factors. The types of modifications are relatively conserved among eukaryotes, however, their effects can vary between kingdoms. In planta, the most studied epigenetic components are primarily DNA methylation of cytosine and histone methylation or acetylation (Pikaard & Mittelsten Scheid, 2014). Furthermore, numerous publications showed that lncRNAs mediate chromatin remodeling by recruiting DNA methyltransferases and histone methyl- and acetyl- transferases (Csorba et al., 2014; Herr, Jensen, Dalmay, & Baulcombe, 2005; Onodera et al., 2005; Pikaard & Mittelsten Scheid, 2014; Y. Wang et al., 2018).

2.1.4.1 DNA Methylation

In the *A. thaliana* genome, the methylation of cytosine has an important role in gene expression control. Cytosine is methylated at the 5th position of the pyrimidine ring, a process mediated by methyltransferases. In *A. thaliana*, there are three known classes of cytosine methyltransferases: 1) DNA (CYTOSINE-5)-METHYLTRANSFERASE 2 (DRM2) which is involved in *de novo* DNA methylation, 2) METHYLTRANSFERASE 1 (MET1), the ortholog of the mammalian DNMT1, primarily responsible for maintaining inherited CpG methylation (Finnegan & Dennis, 1993), and finally 3) the plant specific methyltransferase, DNA (CYTOSINE-5)-METHYLTRANSFERASE 3 (CMT3), which controls the maintenance of CpHpG methylation (X. Liu, Luo, & Wu, 2012). Cytosines in plants are methylated on the same DNA strand in three different variations: CpG, CpHpG, and asymmetric CpHpH (H is A, T, or C). In *A. thaliana* ~18.9% of the nuclear genome contains methylated cytosines with a large majority located at heterochromatic regions (Zhang et al., 2006). CpG methylation predominates with levels of 24%, while CpHpG and CpHpH consist of only 6.7% and 1.7%, respectively (Cokus et al., 2008). Typically, promoter methylation corresponds with gene transcription repression. However, less than 5% of expressed genes' promoters are methylated but the ones that are exhibit lower expression levels and frequently have tissue-specific expression patterns. In mutants of MET1, genes that contain promoter methylation showed less DNA methylation and increased steady-state expression. However, a large majority of genes, approximately one-third, are methylated within the gene body and in *met1* mutant no differences in gene expression were observed. Although gene body methylation is not yet clearly understood, it was found to correlate with genes that are highly transcribed and constitutively expressed (Zhang et al., 2006).

In contrast to mammalian models, *de novo* DNA methylation in plant cells is better understood and shown to be mediated by small RNAi. Regions in the genome found to be associated with siRNAs highly correlated with CpG, CpHpG and CpHpH methylation and are consistent with the known molecular nature of RNA-directed DNA methylation (RdDM) (Cokus et al., 2008). The RdDM pathway is mediated by RNA Pol IV and Pol V where transcripts of RNA Pol IV lead to the production of 24nt-siRNAs that associate with ARGONAUTE 4 (AGO4) and bind to lncRNAs generated by RNA Pol V leading to the recruitment of chromatin remodeling complexes (**Figure 2.3**) (Wierzbicki et al., 2008; Wierzbicki, Ream, Haag, & Pikaard, 2009).

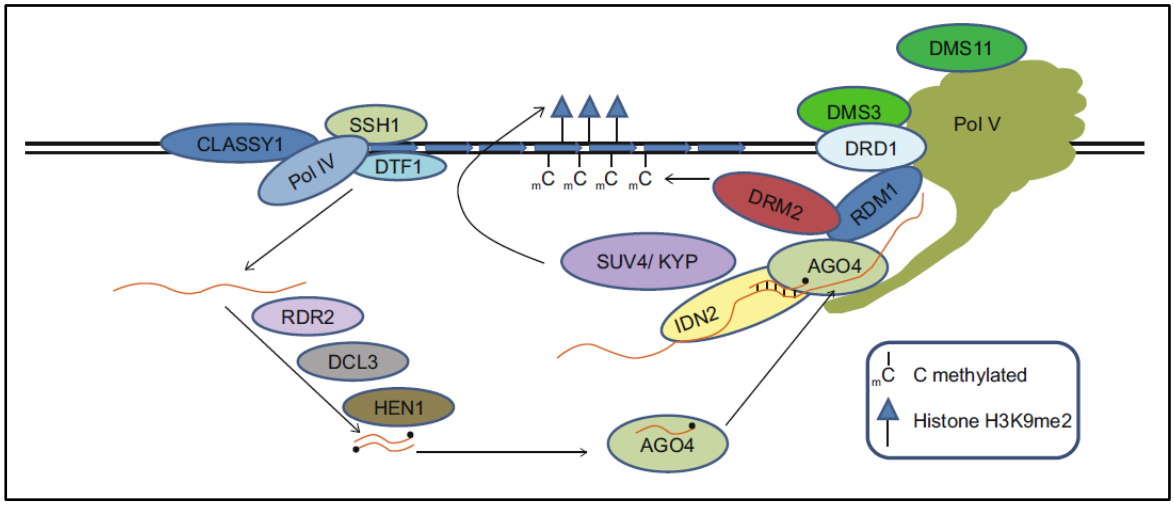


Figure 2.3 Scheme of ncRNA mediated DNA and histone methylation. CLASSY1 (CLSY1), SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) (Law et al., 2013), and DNA-BINDING TRANSCRIPTION FACTOR 1 (DTF1) (J. Liu et al., 2011) regulate the association of RNA Pol IV to the chromatin and ultimately the production of 24nt-siRNAs involved in the RdDM methylation pathway (Greenberg et al., 2011; Zhou, Palanca, & Law, 2018). Single stranded RNA Pol IV RNAs (ssRNAs) are made double stranded by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). Double stranded RNAs (dsRNA) are recognized by endoribonuclease DICER-LIKE 3 (DCL3) which generates 24nt-siRNAs with 2nt overhangs at the 3' ends (Nagano, Fukudome, Hiraguri, Moriyama, & Fukuhara, 2014) and are subsequently methylated by HUA ENHANCER 1 (HEN1) (J. Li, Yang, Yu, Liu, & Chen, 2005). The methylated 24nt-siRNA is then loaded into AGO4 and directs its association with RNA Pol V transcripts modulating DNA and histone methylation via DNA (CYTOSINE-5)-METHYLTRANSFERASE (DRM2) and SET-DOMAIN METHYLTRANSFERASE-SU (VAR) 3–9 HOMOLOGUE 4 (SUV4/KYP), respectively (Cao & Jacobsen, 2002; Jackson, Lindroth, Cao, & Jacobsen, 2002). Figure adapted from (Xu, Tian, & Mo, 2013).

2.1.4.2 Histone Modifications

Histones are commonly known to be "packing" structures in which DNA wraps around and facilitates compaction of the genome in order to fit within the nucleus. Histones also have an important function in transcription regulation. They are highly conserved

amongst eukaryotes and consist of octameric complexes containing two copies each of four core proteins (H2A, H2B, H3, and H4) (Voet et al., 2016). The DNA-histone complex is known as a nucleosome and consists of 147bp of DNA wrapped 1.65 turns around the histone octamer (Luger, Mader, Richmond, Sargent, & Richmond, 1997). Non-nucleosomal DNA, known as linker DNA, ranges between 20-90 bp in length and associates with linker histones (H1 and H5) to further stabilize compaction (Bednar et al., 1998).

Histone core proteins can also be modified via acetylation, methylation, or phosphorylation of lysine and arginine residues at their N-terminal tails affecting the associated DNA (Strahl & Allis, 2000). The modifications can affect chromatin structure in addition to being sites of protein recruitment. The negative charge of the acetyl group (CH3COO-) neutralizes the positive charge of the histone tail liberating the negatively charged DNA allowing increased access of the transcription machinery (Hong, Schroth, Matthews, Yau, & Bradbury, 1993; Lee, Hayes, Pruss, & Wolffe, 1993; Vettese-Dadey et al., 1996). The effects of methylation, on the nucleosome remains unclear, but some studies show that proteins can recognize certain types of methylation and influence gene regulation (Qian et al., 2018). While lysins are observed to be mono-acetylated and associated with transcription activation, the addition of methyl groups offers a more complex regulatory dynamic. Lysins can be mono, di, or tri- methylated which can have positive or negative effects on transcription. Despite this, the effects of some modification types on transcription are well characterized, for example trimethylation of histone 3 lysine 4 or 36 (H3K4me3 or K3K36me3) and acetylation of histone 3 lysine 27 or histone 4 at lysine 5 (H3K27ac or H4K5ac) are associated with transcription activation and referred to as euchromatin modifications, while trimethylation at histone 3 lysine 9 or 27 (H3K9me3 or H3K27me3) leads to silencing and are termed heterochromatin modifications (**Figure 2.4**) (B. Li, Carey, & Workman, 2007). Some known *A. thaliana* methyl- and acetyl- transferases are listed in **Table 2.1**, while arginine modifications are also possible, focus will be placed on the more common and well-studied lysine modifications.

2.1.4.3 Crosstalk between histone and DNA Methylation

Many examples have been described in both mammals and plants showing a correlation between histone modifications and DNA methylation state. For example, H3K9 methylation and hypermethylation of DNA at the same locus are associated with gene silencing (**Figure 2.3**), whereas DNA hypomethylation and the presence of H3K4 methylation are characteristics of active transcription (Tariq & Paszkowski, 2004).

Table 2.1 Histone modifiers in *A. thaliana* (C. Liu, Lu, Cui, & Cao, 2010).

Sites	Proteins involved	Modification
H3K4	ARABIDOPSIS TRITHORX 1/2(ATX1/2)	Methylation
	SET DOMAIN GROUP 4 (SDG4)	Methylation
H3K9	SU(VAR)3–9 HOMOLOGS 4/5/6 (SUVH4/5/6)	Methylation
	SU(VAR)3–9 RELATED 4 (SUVR4)	Methylation
H3K27	ARABIDOPSIS TRITHORAX RELATED 5/6 (ATXR5/6)	Methylation
	CURLY LEAF (CLF)	Methylation
	MEDEA (MEA)	Methylation
	SWINGER (SWN)	Methylation
	CREB BINDING PROTEIN (CBP/p300-like)	Acetylation
H3K36	SET DOMAIN GROUP 4/8/26 (SDG4/8/26)	Methylation
H4K5	HISTONE ACETYLTRANSFERASE OF THE MYST FAMILY 1/2 (HAM1/2)	Acetylation

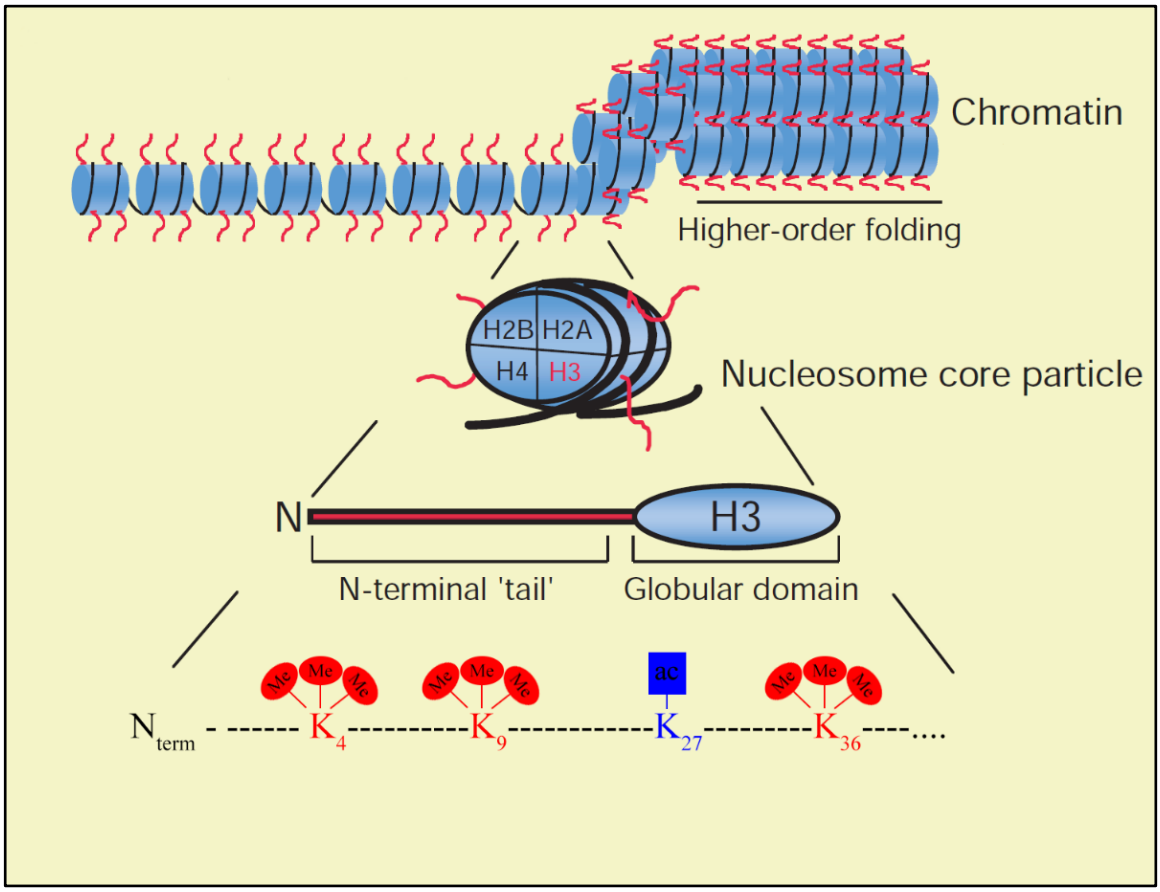


Figure 2.4 Histone modifications. The histone complex is an octamer consisting of two of each core histone protein (H2A, H2B, H3, and H4). Higher-order folding of the chromatin is achieved via the interaction of non-nucleosomal DNA with H1 (not shown). Histone 3 can be modified at lysine residues and occur at the non-globular N-terminal 'tail'. Trimethylation of lysine 4 and 36 are associated with transcription activation while trimethylation of lysine 9 results in silencing. Acylation is typically associated with gene activation and can occur at histone 3 lysine 27 and histone 4 lysine 5 (not shown) Adapted from (Strahl & Allis, 2000).

In addition to methylation, the acetylation state of histones, mediated by HISTONE DEACETYLASES 6 (HDA6), was found to co-localize to the same locus with MET1 and enhance RNA-directed DNA methylation (Aufsatz, Mette, van der Winden, Matzke, & Matzke, 2002; To et al., 2011). Furthermore, histone methyl transferases SUVH4, SUVH2, and SUVH9 contain the SET and RING-associated (SRA) domain which allows for binding to methylated DNA. While they resemble histone methyltransferases, SUVH2 and SUVH9 lack methyltransferase activity, but retained their ability to bind to methylated DNA and mediate gene silencing via RdDM by recruiting RNA Pol V (Johnson et al., 2014). In contrast, SUVH4 retains its methyltransferase activity, and after binding to methylated DNA, is able to mediate H3K9me and H3K27me (Lindroth et al., 2004). Furthermore, when histone 3 is concurrently methylated at lysine 9 and 27, CMT3 can directly interact with N-terminal tail of the histone leading to CpHpG methylation; suggesting a sort of a 'histone code' for the recruitment of CMT3 to methylate the DNA loci (He, Chen, & Zhu, 2011)

2.2 *Uridine diphosphate glycosyltransferases (UGTs)*

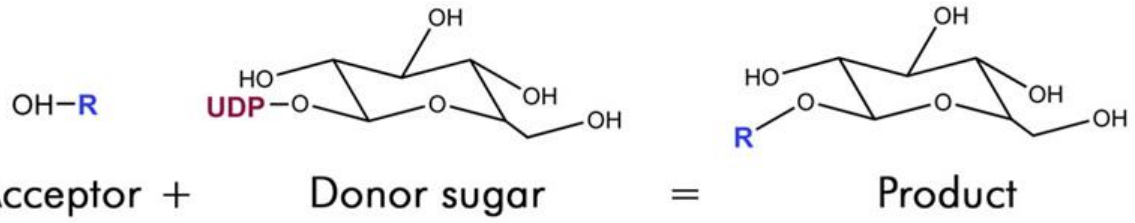
The covalent addition of sugars to molecules can have profound effects on their bioactivity, solubility, and transport properties within the cell and throughout the plant (Rademacher, Parekh, & Dwek, 1988; Ross, Li, Lim, & Bowles, 2001). The transfer is catalyzed by uridine diphosphate glycosyltransferases (UGTs) which add sugars to their target substrates, primarily at hydroxyl, carboxyl, or amine functional groups, using uridine diphosphate activated sugars (UDP-sugar) as a donor (**Figure 2.5A**) (Meech et al., 2019). Plant UGTs are characterized by the plant secondary product glycosyltransferase (PSPG) motif, a consensus sequence of 44 amino acids near the C-terminus, that is responsible for the transferase activity (**Figure 2.5B**) (Hughes & Hughes, 1994). Structural analysis revealed highly conversed histidine (position 19) and glutamic acid (position 27) residues in the PSPG motif deprotonates the acceptor hydroxyl group leading to a nucleophilic attack at the C1 of the sugar donor to form a glycosidic bond (Kapitonov & Yu, 1999).

In higher plants, members of the UGT superfamily glycosylate a broad range of substrates, including plant hormones, all major classes of plant secondary metabolites, and xenobiotics such as herbicides. UGTs regulates many properties of the aglycones. In *A. thaliana*, there are 107 putative UGT genes classified into 14 groups (A–N) according to their sequence similarity, while many remain uncharacterized, some have been found to play important roles in plant development and defense (Paquette, Moller, & Bak, 2003; Ross et al., 2001). Brassinosteroid, for example, is an important plant hormone involved in plant

growth and senescence and is glycosylated by UGT73C5 and UGT73C6 rendering the molecule biologically inactive. *A. thaliana* plants overexpressing these genes display cabbage like growth characteristics of brassinosteroid deficient mutants (Husar et al., 2011; Noguchi et al., 1999; Poppenberger et al., 2005). Moreover, UGT73C5 was shown to inactivate the *Fusarium* mycotoxin deoxynivalenol (DON) and overexpression in *A. thaliana* led to greater tolerance when plants were grown on plates (Poppenberger et al., 2003). Similarly, homeostasis of the crucial hormone indole-3-acetic acid (IAA) is regulated in part by *UGT84B1* and when overexpressed leads to an accumulation of its product 1-*O*-(indol-3-acetyl)-*β*-D-Gluc (R. G. Jackson et al., 2002). In addition, multiple UGTs are differentially regulated during abiotic or biotic stresses and shown to influence tolerance and resistance. Salicylic acid (SA) is a phytohormone involved in pathogen defense and is glycosylated by UGT74F1 and UGT74F2 (Lim et al., 2002). While the effects of misregulation on defense have yet to be studied, each enzyme uses a specific hydroxyl group as an acceptor for the formation of the glycosidic bond yielding salicylic acid glucoside and salicylic acid glucose ester, respectively. The additions are thought to affect how SA is stored, further metabolized, and contributes to defense response (J. V. Dean & Delaney, 2008; George Thompson, Iancu, Neet, Dean, & Choe, 2017). In contrast, even though *UGT73B3* and *UGT73B5* are induced by SA the substrates remain unknown, however, when knocked out have a decreased resistance to the bacteria *Pseudomonas syringae* pv *tomato* (Langlois-Meurinne, Gachon, & Saindrenan, 2005).

While many UGTs have activity to certain molecules yielding specific products, they remain promiscuous due to their ability to glycosylate more than one substrate. Flavonols are commonly glycosylated by various UGTs in *A. thaliana* further underling the extent of complexity and redundancy of these enzymes (Lim, Ashford, Hou, Jackson, & Bowles, 2004). Nonetheless, UGTs in the same group have been shown to have similar substrates. Members of groups F and O were found to selectively glycosylate flavonols and zeatin, respectively. However, this work focuses on *UGT76E12* and *UGT76E11*, both found in group H which members have been shown to be involved in different biological processes including plant defense and cytokinin homeostasis (Brazier-Hicks, Gershater, Dixon, & Edwards, 2018). Recent publications have shown some evidence that *UGT76E12* and *UGT76E11* may be involved in salt stress response and jasmonic acid (JA) modifications (Haroth et al., 2019; Q. Li et al., 2018). Furthermore, *UGT76E12* and *UGT76E11* are overlapped by two lncNATs, that we *NAT-UGT76E12* and *NAT-UGT76E11*, respectively (**Figure 2.6**).

A



B

Figure 2.5 ***A***, Transfer of UDP-glucose to an aglycon producing a glucoside product (http://www.p450.kvl.dk/UGT.shtml). ***B,*** Plant secondary product glycosyltransferase (PSPG) motif conserved in plant family 1 UGTs found near the C-terminus. Histidine 19 glutamic acid 27 are crucial for the glycosyltransferase activity (Osmani, Bak, & Moller, 2009).

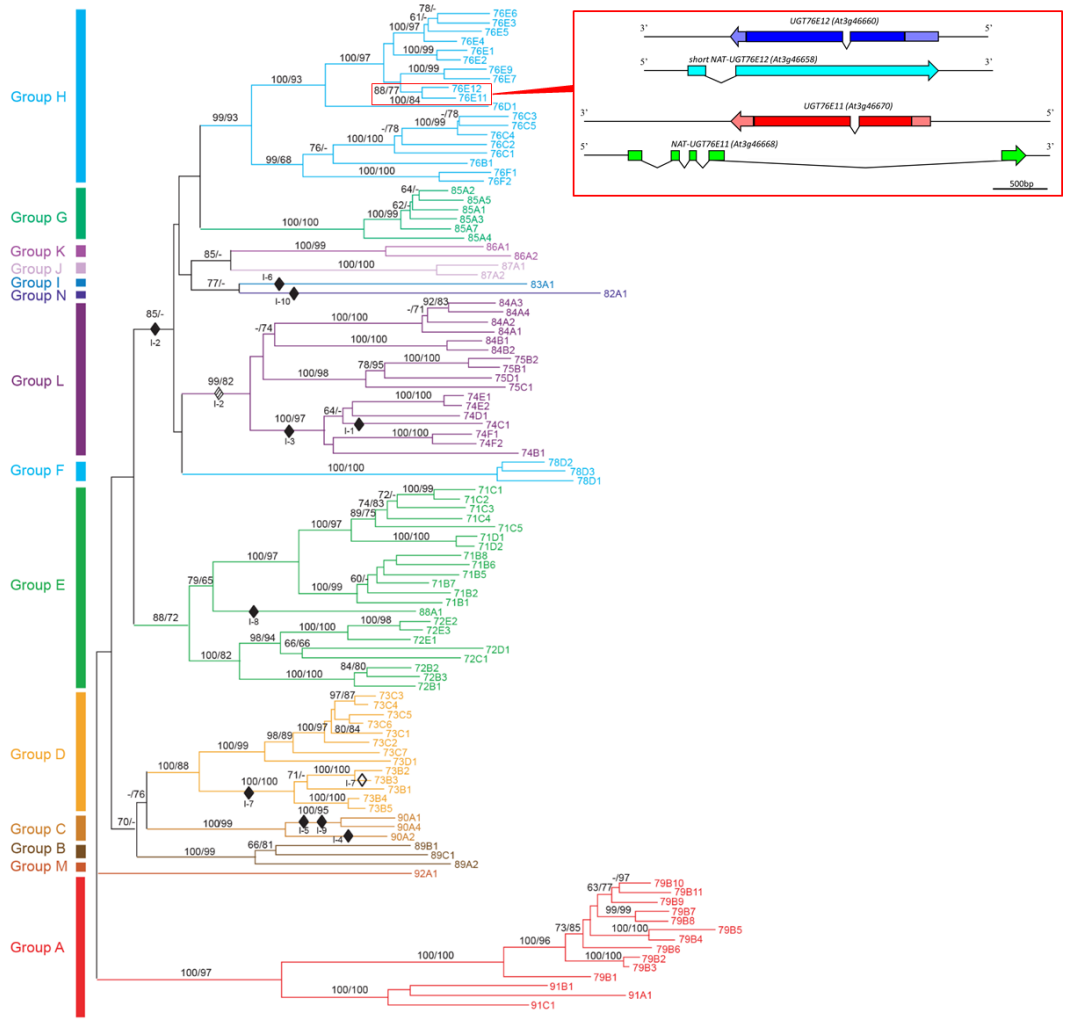


Figure 2.6 UGT Superfamily. *A. thaliana* contains 107 UGT genes classified into 14 groups (A–N) according to their sequence similarity. The 76E subfamily is located in group H (light blue). *76- E3*, *E4*, *E5*, *E6*, *E11*, and *E12* are found near the same locus. Adapted from (Ross et al., 2001) Boxed in red are *UGT76E12* and *UGT76E11* and their overlapping lncNATs.

2.3 *Abiotic and biotic stress*

In contrast to a controlled laboratory setting, plants must maintain a balance and survive a multitude of concurrent environmental challenges. Simultaneous exposure to abiotic and biotic stresses is rarely studied largely due to the substantial amount of crosstalk between the two responses (**Figure 2.7**). Plant hormones in particular, have crucial functions in the response to a wide vary of stresses (Bari & Jones, 2009). Abiotic stresses including drought, salt, cold, and heat mainly activates the abscisic acid (ABA) pathway while jasmonic acid (JA), salicylic acid (SA), and ethylene (ET) are mainly involved in biotic stresses (Bari & Jones, 2009; Verma, Ravindran, & Kumar, 2016). The crosstalk between ABA and JA pathways is mediated by DELLA proteins (Verma et al., 2016). ABA enhances the expression of the DELLA protein *RGA-LIKE2* (*RGL2*), which binds the JA ZIM-domain 1 (JAZ1) repressor, preventing JAZ1's inhibitory effect on JA responsive genes (Hou, Lee, Xia, Yan, & Yu, 2010; Piskurewicz et al., 2008). Furthermore, gibberellic acid (GA), a plant hormone involved in promoting multiple aspects of plant growth and development including seed germination, flowering, and senescence, suppresses expression of JA responsive genes by triggering the degradation of DELLA proteins (Daviere & Achard, 2013; Piskurewicz et al., 2008). In low GA levels, ABA levels increase which results in an accumulation of DELLAs and as a result an increased expression of JA responsive gene. While the crosstalk is vastly more extensive and complex, DELLAs link ABA and JA responses and gives insight on how plants deal with biotic, specifically necrotrophic infections, and abiotic stresses. In addition, their relationship with GA levels underlines a connection with plant growth and development. (Piskurewicz et al., 2008).

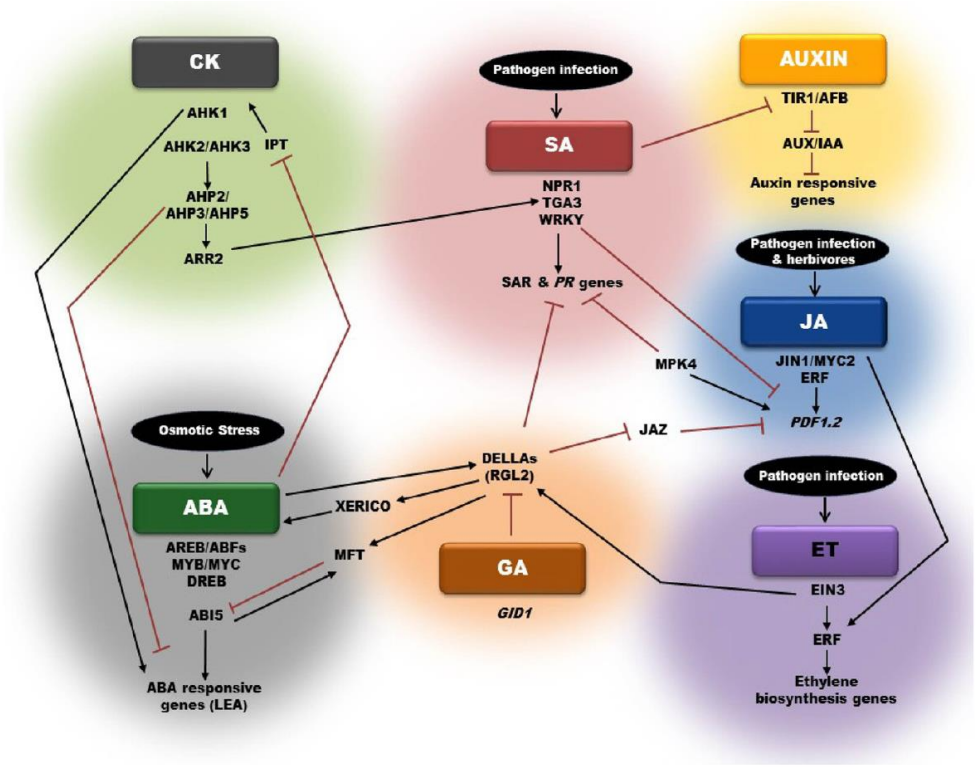


Figure 2.7 General overview of crosstalk between major plant hormones. JA and ET are the main effectors in response to necrotrophic pathogens while SA is mainly involved in autotrophic infections. JA and ET function in tandem in regulating defense responsive genes. DELLA proteins are key in the crosstalk between ABA and JA signaling. Overview from (Verma et al., 2016)

2.4 *B. cinerea*

Considered the second most detrimental phytopathogen worldwide, *B. cinerea* is a fungus that infects more than 500 plant species (R. Dean et al., 2012; Elad, Pertot, Prado, & Stewart, 2016). While primarily infecting dicotyledonous hosts it can also infect a variety of monocots (Williamson, Tudzynski, Tudzynski, & van Kan, 2007). It infects almost all cash crops including vegetables, fruits, and flowers leading to annual global losses between $10 billion to $100 billion (Elad et al., 2016; Weiberg et al., 2013). This degree of damage is largely in part due to its ability to cause damage pre- and post-harvest, affect various tissue types, remain dormant until optimal conditions, and increasing resistance to fungicides (De Miccolis Angelini et al., 2014; Elad et al., 2016).

2.4.1.1 Pathogenesis

B. cinerea is more commonly known as grey mold disease due to the color of the conidia that develops during an infection. It is a necrotrophic fungus requiring death of the host cells in order to grow and propagate which results in the formation of large necrotic lesions (Williamson et al., 2007). Progression of the infection under optimal conditions can be quite rapid and occurs in a matter of days, however, it can also enter a state of dormancy

in suboptimal conditions. The infection starts when *B. cinerea* encounters the host and releases cutinase and lipase enzymes to breach the cuticle layer. Once passed the cuticle, pectinases and cellulases are secreted to further degrade underlying epidermal cells. After penetration of the outer protective layers, *B. cinerea* releases various phytotoxic metabolites and proteins to induce cell death. Botrydial, the most studied metabolite produced by the pathogen, greatly affects the virulence of some strains which require it to kill host cells (**Figure 2.8**) (Williamson et al., 2007). Furthermore, a study in *A. thaliana* revealed that small RNAs secreted by *B. cinerea* utilize the host RNA interference (RNAi) machinery by binding to ARGONAUTE 1 (AGO1) and selectively silencing host immunity genes (Weiberg et al., 2013). Taken together, all the evolved infection strategies allow *B. cinerea* to affect hundreds of plant species and make it a formidable agricultural foe.

2.4.1.2 Host Defense

While *B. cinerea* has evolved numerous ways to infect their hosts, plants have also adopted several defense strategies. When infected, many protective immune signaling pathways are activated including pathogen-derived microbial associated molecular patterns (MAMPs), host damage-associated molecular patterns (DAMPs), and the plant defense hormones JA and ET. In addition, crosstalk between each pathway results in a local and systemic immune response (AbuQamar, Moustafa, & Tran, 2017). Generally, immune defense against biotrophic and, in part, hemi-biotrophic pathogens is mediated by SA. However, necrotrophic and late hemi-biotrophic infections result in a JA and ET response (AbuQamar et al., 2017). Additionally, these hormones are antagonists of each other, and it has been shown that SA represses the JA pathway by reducing the accumulation of OCTADECANOID-RESPONSIVE ARABIDOPSIS AP2/ERF59 (ORA59), a transcription factor (TF) that binds to GCC-box motifs found in the promoters of JA responsive genes (Van der Does et al., 2013). Remarkably, *B. cinerea* has taken advantage of this relationship and manipulates it by producing exopolysaccharide, an elicitor of the SA pathway, resulting in decreased expression JA responsive genes (El Oirdi et al., 2011). The importance of JA and ET in *B. cinerea* infection is further emphasized in the cellulose synthase *CeSA3* mutant (*cev1*) where their production is increased and as a result is insensitive to the antagonistic effect of SA (Ellis, Karafyllidis, Wasternack, & Turner, 2002). Consequently, JA responsive genes including *PLANT DEFENSIN 1.2* (*PDF1.2*) and *VEGETATIVE STORAGE PROTEIN 2* (*VSP2*) are up regulated, conferring resistance to *B. cinerea* infection (Leon-Reyes et al., 2010).

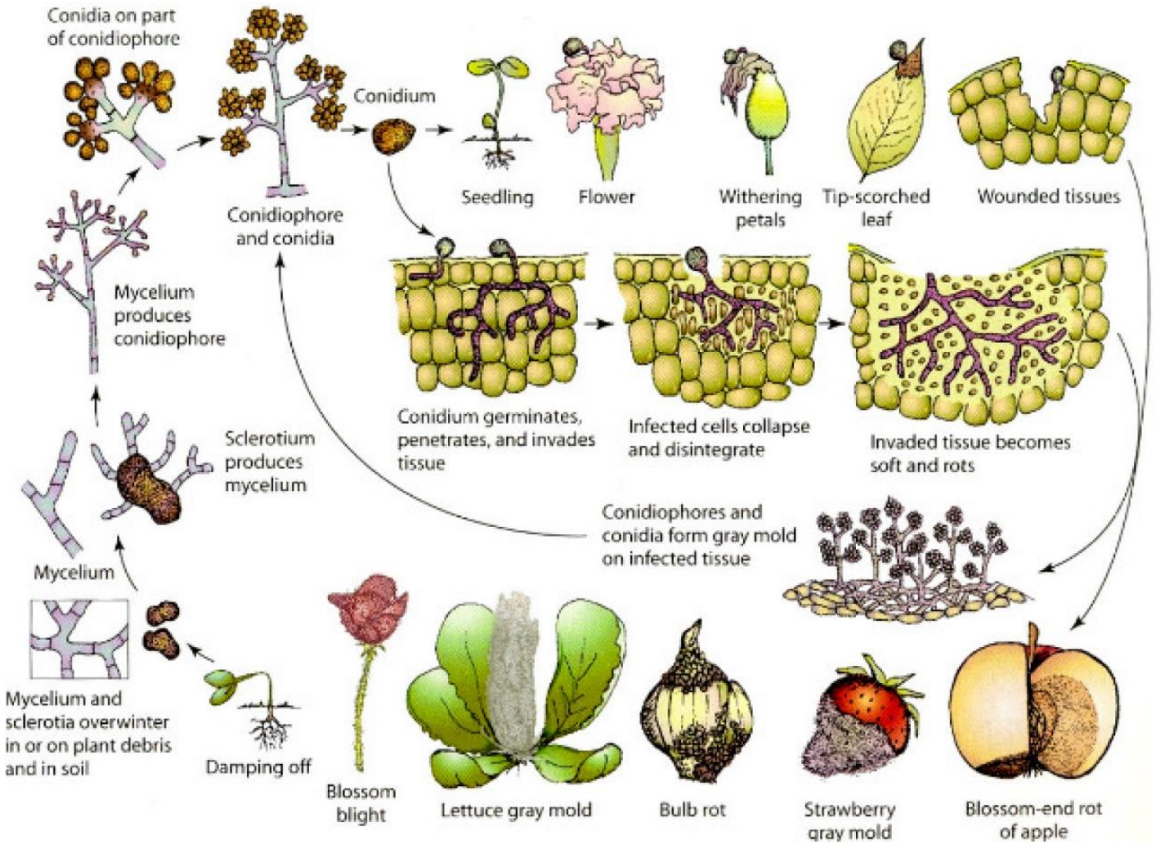


Figure 2.8 *B. cinerea infection*. Life cycle progresses in the following stages: penetration, necrosis of host tissue, and sporulation. The life cycle can begin from either sclerotium or conidium. Sclerotium is a dormant state that is comprised of compacted mycelium. Once conditions become favorable for growth, the sclerotium germinate and produce mycelium leading to the creation of conidiophore and conidia. Conidia in comparison to scleortium are short-lived but are the inoculating component of *B. cinerea*. Germination of conidium begins when they encounter the host. Mycelium grow from the conidium and release a multitude of necrotrophic enzymes and metabolites leading to disintegration and maceration of the tissue. As the infection progresses lesions begin to form and eventually leads to the growth of conidiophore and subsequently more conidium. *B. cinerea* can infect multiple species and tissue types including: flowers, leaves, root structures, and fruit. (Agrios, 2005).

3 Thesis Objectives

Previously considered "transcriptional noise," long non-coding RNAs (lncRNAs) have been discovered to have important roles in a multitude of molecular mechanisms. Their functions in mammals have been extensively studied and shown to regulate important developmental process like chromosome X-inactivation (Monfort & Wutz, 2017). In addition, an increasing amount of data further highlights their importance linking them to various cancers (Bach & Lee, 2018). Likewise, a growing amount of research in planta is revealing their significance in the regulation of several developmental processes and in response to stress conditions (H. V. Wang & Chekanova, 2017). However, in comparison to mammalian models, lncRNAs in plants remain poorly understood. This thesis work focuses on long non-coding natural antisense transcripts (lncNATs), a lncRNA subtype, and their potential role in gene expression regulation. As mentioned in the previous section, only a handful of lncNATs have been described and the mechanism of only one has been almost completely elucidated. Moreover, most of the studied lncNATs overlap single genes and less is known about ones associated with members of multigene families. Therefore, the purpose of this work is to characterize a lncNAT in *A. thaliana* and to evaluate its mechanism of action in the context of a multigene family. Here we focus on *NAT-UGT76E12*, the natural antisense transcript associated with *UGT76E12*, a gene belonging to the UDP glycosyltransferase (UGT) superfamily. In addition, we examine *NAT-UGT76E11*, a natural antisense transcript overlapping *UGT76E11* a gene with 89% sequence percent identity and located adjacent to *UGT76E12*. Using these genes, we aim to answer the following questions:

1) Do *NAT-UGT76E11* and *NAT-UGT76E12* regulate the expression of their sense genes?

2) Is the regulation exclusive to the sense gene?

- Could they influence genes with high sequence similarities?

3) Do they function in *cis* and/or *trans*?

- Does the regulation happen at the gene locus or elsewhere?

4) What is the mechanism of action?

- Do they have a positive or negative effect on expression?

Additionally, since the roles of two enzymes encoded by the protein-coding genes are not known, we also aim to investigate their potential function. With this work we hope to further punctuate the importance lncNATs in plants and add insight to their mechanism of action in gene regulation control.

4 Results

Part I – Analysis of *NAT-UGT76E12*

4.1 *Three independent natural antisense transcripts overlap the UT76E11 and UGT76E12 genes*

In order to validate transcript information from the TAIR database (The Arabidopsis Information Resource), the transcription initiation and termination sites, splicing pattern, and expression levels of the sense and antisense genes were analyzed. Using RACE, we found the transcription termination site (TTS) and transcription initiation site (TIS) of *NAT-UGT76E12* ranged between 0 to 245 nt longer and 33 nt shorter than the TAIR annotation, respectively (**Supplementary Figure 8.1**). Data from a tiling array database reported by Matsui and collaborators suggested the presence of a 13.8kB transcript overlapping multiple UGTs including *UGT76E11* and *UGT76E12* (**Supplementary Figure 8.2**) (Matsui et al., 2010). We were able to detect the presence of a longer variant of *NAT-UGT76E12*, however, much shorter than the tiling array predictions with a transcription initiation site starting at the 5' end of *NAT-UGT76E12* spanning to the 3' end of *NAT-UGT76E11* annotated in TAIR (**Figure 4.1**). Hereafter, the short and long variants of *NAT-UGT76E12* will be indicated as *shNAT-UGT76E12* and *lgNAT-UGT76E12,* respectively. In addition, we also chose to include in our analysis a third antisense transcript overlapping *UGT76E11*, *NAT-UGT76E11*. Next, we examined the splicing pattern and found, after sequencing analysis that all the annotated introns were spliced (**Figure 4.2A**). The splicing of the largest intron shared by *NAT-UGT76E11* and *lgNAT-UGT76E12* results in a mature transcript without sequence complementarity to *UGT76E11* suggesting they may not play a role in its regulation. When examining the expression levels, we found that *NAT-UGT76E11* was higher than *shNAT-UGT76E12* and that *lgNAT-UGT76E12* was considerably lower (**Figure 4.2B**). With this data we were able to conclude the presence of three independent natural antisense transcripts *shNAT-UGT76E12*, *lgNAT-UGT76E12*, and *NAT-UGT76E11* with maximum estimated genomic sizes of 2756, 6447, and 3618 bp, respectively. In all cases complete intron splicing was observed yielding maximum final transcripts sizes of 2484, 3183, and 626 bp, respectively.

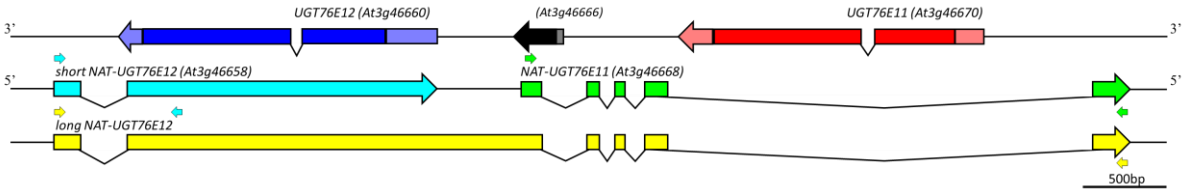


Figure 4.1 Layout of the genomic region comprising *UGT76E12* and *UGT76E11* (TAIR10). The long variant of the natural antisense transcript (*lgNAT-UGT76E12*) starts at the 5' end of *NAT-UGT76E12* and spans to the 3' end of *NAT-UGT76E11*. *shNAT-UGT76E12:* (unspliced 2756bp / spliced 2484bp), *lgNAT-UGT76E12:* (unspliced 6447bp / spliced 3183bp), *NAT-UGT76E11:* (unspliced 3618bp / spliced 626bp). Size determined by sequencing and experimental data. Small arrows indicate position of primers used for RT-PCR in **Figure 4.2**.

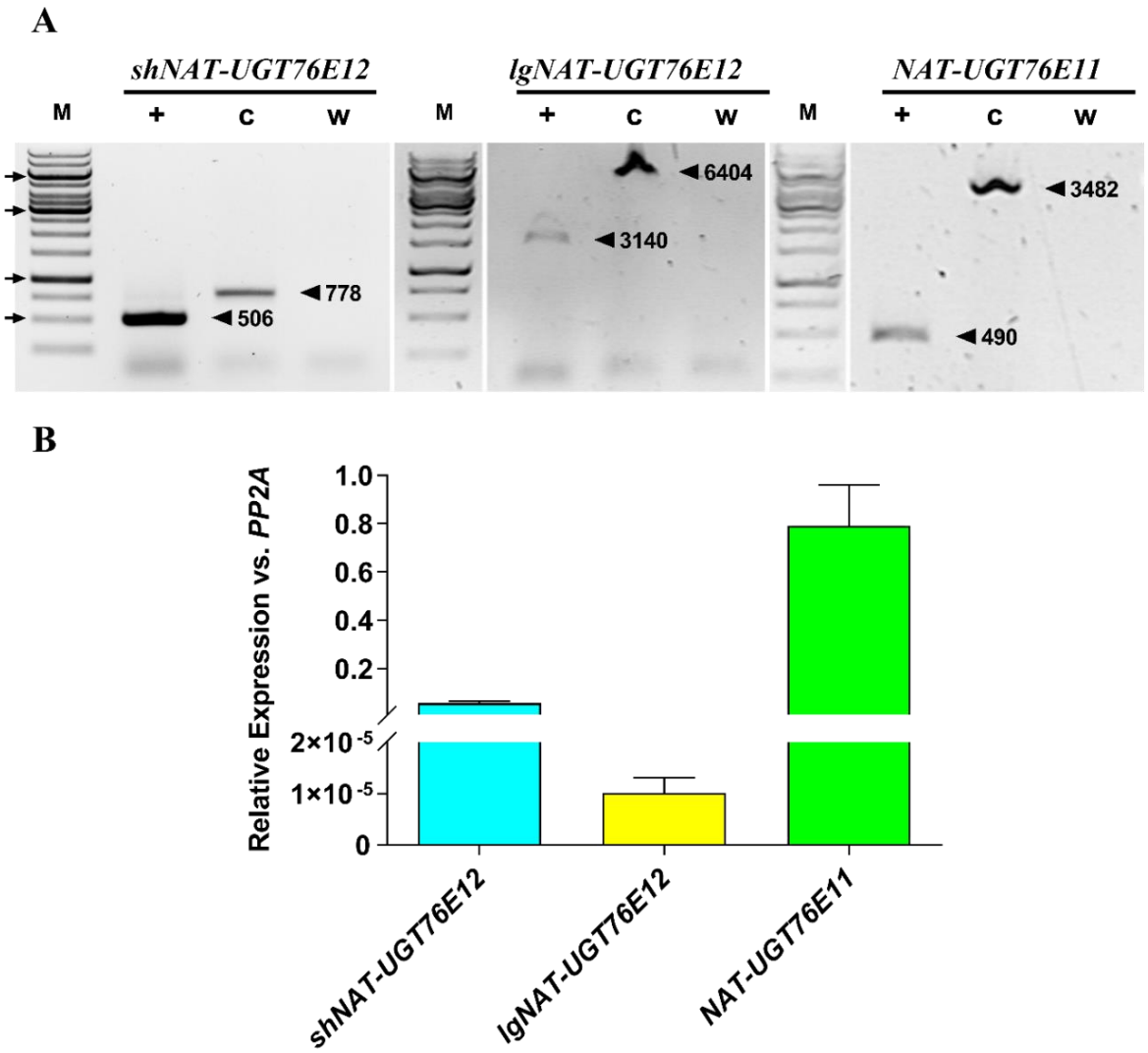


Figure 4.2 Splicing pattern and expression analysis of *shNAT-UGT76E12*, *lgNAT-UGT76E12*, and *NAT-UGT76E11*. *A*, RT-PCR showing spliced mature RNA. Marker, M; cDNA reactions with reverse transcriptase, +; genomic control, c; water control, w. The arrowhead indicates the PCR product and the size is shown on the right. Arrows on molecular weight marker indicate 0.5 kb, 1 kb, 3 kb, and 6 kb (bottom to top). Primers used are indicated **Figure 4.1**. ***B*,** Expression levels in 10 day old seedlings determined by qRT-PCR. Data are mean ± SD from 1 representative experiment with 3 biological replicates n=3.

4.2 *NAT-UGT76E12 expression is higher than UGT76E12 at different developmental stages*

Next, we aimed to investigate the relationship between the expression of the lncNATs and their overlapping sense genes. From a microarray database (Genevestigator), where the expression across various developmental stages was investigated, we observed changes in the expression of both *UGT76E11* and *UGT76E12* during development (Hruz et al., 2008). In addition, a tiling array from Matsui and collaborators suggested an anticorrelated relationship between *NAT-UGT76E12* and *UGT76E12* under several stress conditions (Matsui et al., 2010). Nonetheless, many databases contain limited information about non-coding RNAs, so we investigated how the changes in expression of *UGT76E11* and *UGT76E12* correlate with their respective NATs at difference time points. The term *NAT-UGT76E12* will refer to levels of both *shNAT-UGT76E12* and *lgNAT-UGT76E12* combined as the expression of *lgNAT-UGT76E12* alone was below the reliable detectable threshold in the rosettes across all time points under the experimental conditions. Initially examining rosettes from 14 day old plants, we found the basal expression of *UGT76E12* to be substantially less than *NAT-UGT76E12.* In contrast, the expression of *UGT76E11* was higher than that of *NAT-UGT76E11* at earlier time points (**Figure 4.3A**). Under our experimental conditions bolting occurred at 25 days, therefore, we collected rosettes at 22 and 25 days to examine differences before and after flowering initiation. Our results indicated that the event of flowering initiation had no obvious impact on the overall expression level dynamic of the analyzed genes. At later time points, the expression dynamic between *UGT76E12* and *NAT-UGT76E12* remained constant whereas after 35 days the expression of *NAT-UGT76E11* exceeded *UGT76E11* (**Figure 4.3B**). In summary, we observed a simultaneous increase across different stages of senescence between *UGT76E12* and *NAT-UGT76E12* suggesting *NAT-UGT76E12* may have a positive influence on the expression of its sense gene. In contrast, *UGT76E11* and *NAT-UGT76E11* seem to have an anti-correlated relationship, whereas the expression of *NAT-UGT76E11* increases the expression of *UGT76E11* decreases as the plant ages (**Figure 4.3B**).

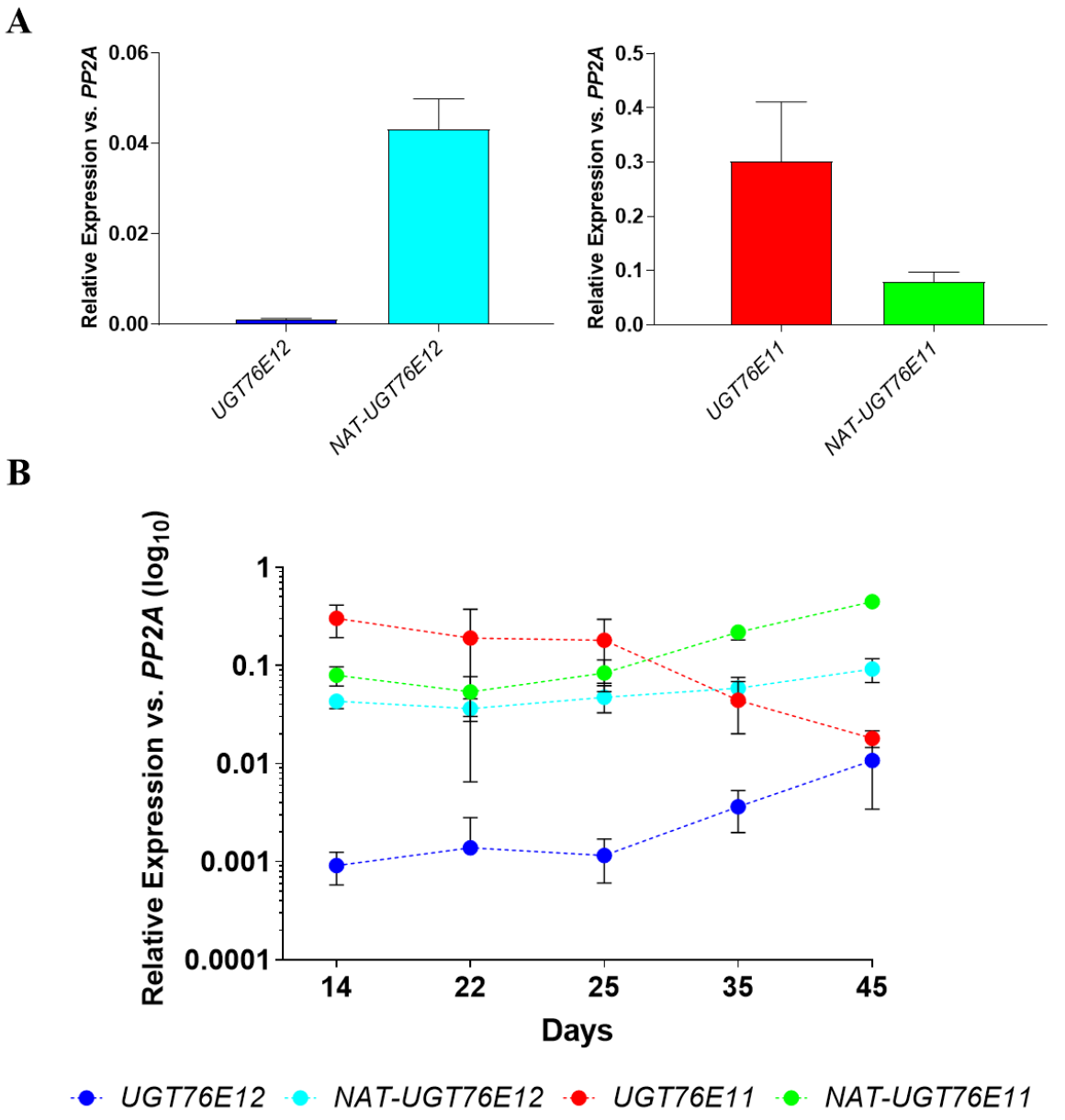


Figure 4.3 Expression of *NAT-UGT76E11* and *NAT-UGT76E12* and their sense genes. *A,* qRT-PCR over total RNA extracted from whole rosettes from 14 day old plants grown on soil. ***B,*** qRT-PCR of samples from plants at various stages of development To observe changes before and after flowering initiation samples were collected at 22 and 25 days, respectively. Data are mean ± SD from three independent experiments. Each sample contains pools of 4-12 rosettes with removed stems.

4.3 *NAT-UGT76E12 and UGT76E12 are expressed in roots.*

Co-expression of *NAT-UGT76E12* and *UGT76E12* in the same tissue would be a requirement for a positive correlation. To test this question, we examined promoter activity using *β-Glucuronidase* (*GUS*) as the reporter gene and expression levels in shoots and roots. In order to ensure a complete capture of the theoretical promoter region, we generated two independent reporter gene lines for *NAT-UGT76E12* using 1057 and 2350 bp (simplified 1kB and 2kB) upstream of the transcription initiation site (**Supplementary Figure 8.3A**).

Since three independent lines from *NAT-UGT76E12promoter$_{1kB}$::GUS* and *NAT-UGT76E12promoter$_{2kB}$::GUS* constructs showed identical promoter activity (not shown), we focused on the *NAT-UGT76E12promoter$_{1kB}$::GUS* construct as an output for *NAT-UGT76E12* promoter activity. In addition, reporter gene lines for *UGT76E11* and *UGT76E12* were also generated (**Supplementary Figure 8.3B**). Several lines were analyzed and three independent lines for each construct sharing a similar pattern of promoter activity at 10 and at 14 day after stratification were selected (**Supplementary Figure 8.4**, **Supplementary Figure 8.5**). One representative line from each construct was then used for the time course experiment. Plants for time points after 25 days were grown on soil to prevent a stress induced response. We observed high promoter activity of *UGT7E12* throughout the seedling at 3 days and a gradual decrease at later time points. Interestingly, only at 10 days does the promoter of *NAT-UGT76E12* begin to show activity and when compared to *UGT76E12* we see at the same and subsequent time points that both promoters are solely active in the root tissue (**Supplementary Figure 8.6**, **Supplementary Figure 8.7**). Furthermore, the promoter of *UGT76E11* was observed to mainly be active in leaves across all time points (**Supplementary Figure 8.8**). We also examined flowers and found the *UGT76E12* promoter to be active in the stamen and the *UGT76E11* promoter active in the carpel and siliques. Next, we examined the expression levels and surprisingly, in contradiction to the reporter gene lines, found both *NAT-UGT76E12* and *UGT76E12* were detected at the same levels in the shoot and root tissue. Moreover, when quantifying levels of *UGT76E11* transcript, we find a strong correlation with the reporter gene lines and only see expression in the shoot (**Figure 4.4**). We additionally found *NAT-UGT76E11* to be expressed throughout the seedling (**Supplementary Figure 8.9**). Taken together, we observed that the promoters of *NAT-UGT76E12* and *UGT76E12* after 10 days are both active in roots, however, *UGT76E12* can be expressed independent of *NAT-UGT76E12* at earlier stages of development. The discrepancy between the reporter gene lines and expression data for *NAT-UGT76E12* and *UGT76E12* can be explained by the relatively low sensitivity of the histochemical analysis, especially regarding promoters with low activity.

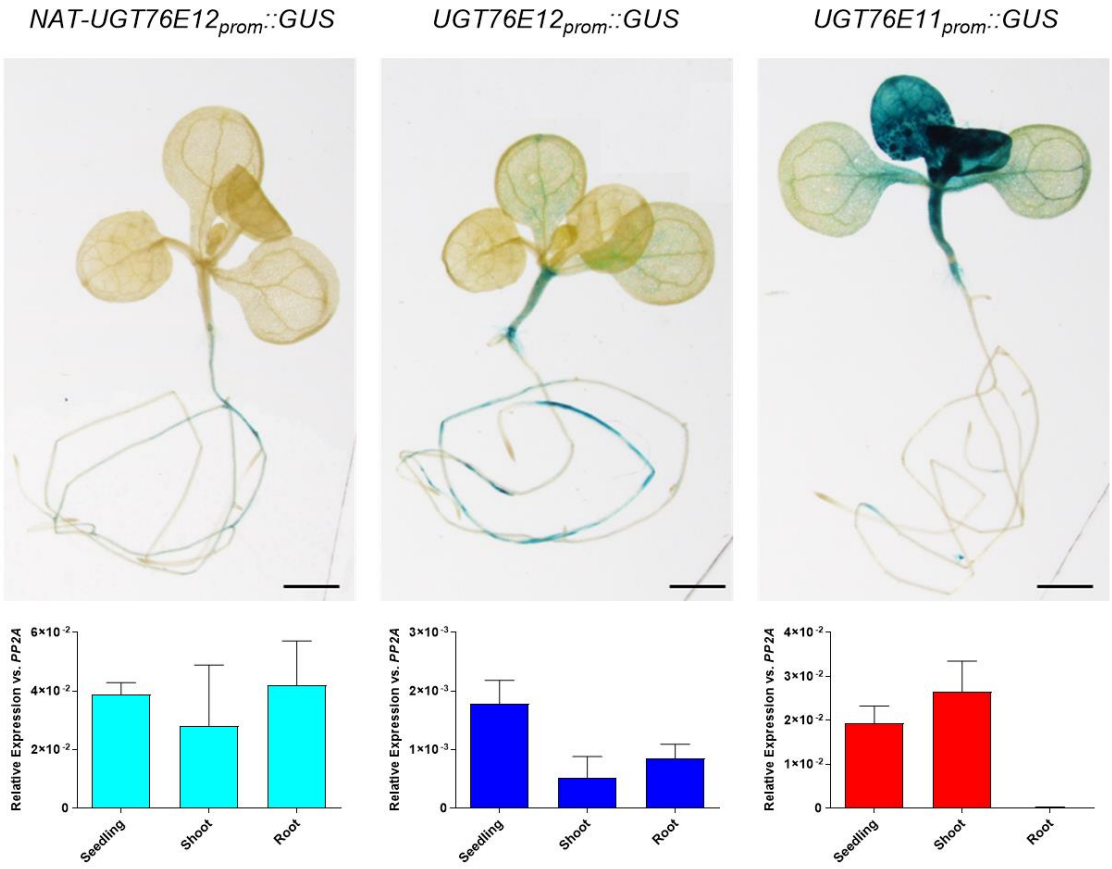


Figure 4.4 Promoter activity and expression levels in shoots and roots. ***Top,*** Promotor activity in 10 day old seedlings of representative transgenic reporter lines grown on plates. Staining was performed overnight at 37ºC. Scale 1.0 cm. ***Bottom,*** qRT-PCR comparing expression in shoots and roots of 10 day old seedlings. Data are mean ± SD from 3 biological replicates from 1 of 3 independent experiments with similar results.

4.4 *Increased NAT-UGT76E12 expression in cis but not trans affects UGT76E12 expression*

In order to further elucidate how the expression of *shNAT-UGT76E12* and lg*NAT-UGT76E12* influences *UGT76E12* levels, we independently overexpressed the genomic sequence of both variants (**Figure 4.5A**). Of the 5 independent *shNAT-UGT76E12ox* lines, 3 were chosen to be analyzed based on levels of overexpression ranging from low, medium to high (**Figure 4.5B**). Due to a high frequency of silencing of the construct in *A. thaliana*, only 2 lines overexpressing *lgNAT-UGT76E12* were obtained (**Figure 4.5D**). When analyzing *UGT76E11* and *UGT76E12* in these lines we detected no significant differences (**Figure 4.5C, E**). It is important to note that the level of *UGT76E12* in 10 day old seedlings grown in vertical plates nears the border of reliable detection making its measurement difficult. Nonetheless, we concluded that ectopic overexpression of *shNAT-UGT76E12* or *lgNAT-UGT76E12* does not obviously influence *UGT76E11* or *UGT76E12* expression in a

trans manner. When analyzing *NAT-UGT76E11* expression levels in the *shNAT-UGT76E12ox* lines they were only marginally affected providing additional evidence of the presence of an autonomous independent promotor (**Figure 4.5C**). In contrast, *NAT-UGT76E11* levels in the *lgNAT-UGT76E12ox* lines were increased to a greater extent.

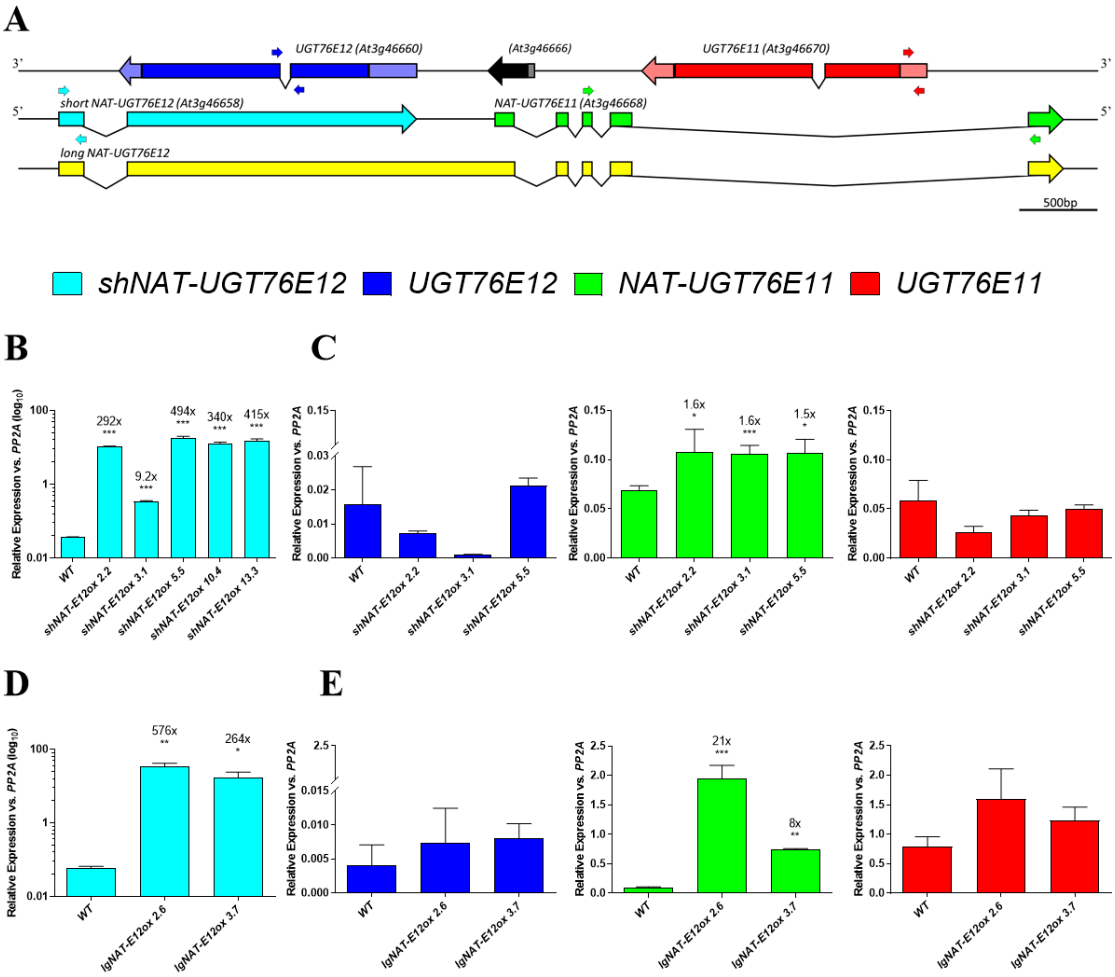


Figure 4.5 Effect of *shNAT-UGT76E12* and *lgNAT-UGT76E12* overexpression. *A,* Genomic layout showing location of primers used in qRT-PCR (the color of the small arrows corresponds to the target gene). ***B, D,*** qRT-PCR showing the level of *NAT-UGT76E12* overexpression in the *shNAT-UGT76E12ox* (***B***) and *lgNAT-UGT76E12ox* (***D***) lines. ***C, E,*** qRT-PCR showing levels expression of sense and antisense genes in selected *shNAT-UGT76E12ox* (***C***) and *lgNAT-UGT76E12ox* (***E***) lines. Total RNA was extracted from 10 day old seedlings. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. Two-tailed t-test; *p<0.05, **p<0.01, ***p<0.001

Since the *lgNAT-UGT76E12ox* lines contain the sequences of all three NATs, we compared the expression levels of each transcript in these lines. The qRT-PCR primers that target *shNAT-UGT76E12* and *NAT-UGT76E11* overlap *lgNAT-UGT76E12* and in order to obtain their true values, the expression level of *lgNAT-UGT76E12* was measured using primers targeting a non-overlapping region (**Figure 4.6A**) and the values subtracted from

the levels of *shNAT-UGT76E12* and *NAT-UGT76E11* (**Figure 4.6B**). Although, the degree of *lgNAT-UGT76E12* overexpression was substantially higher compared to the wild type, we observed that a large majority of the transcripts stop at the 3' end of *shNAT-UGT76E12* indicating the presence of transcriptional terminator (**Figure 4.6C**).

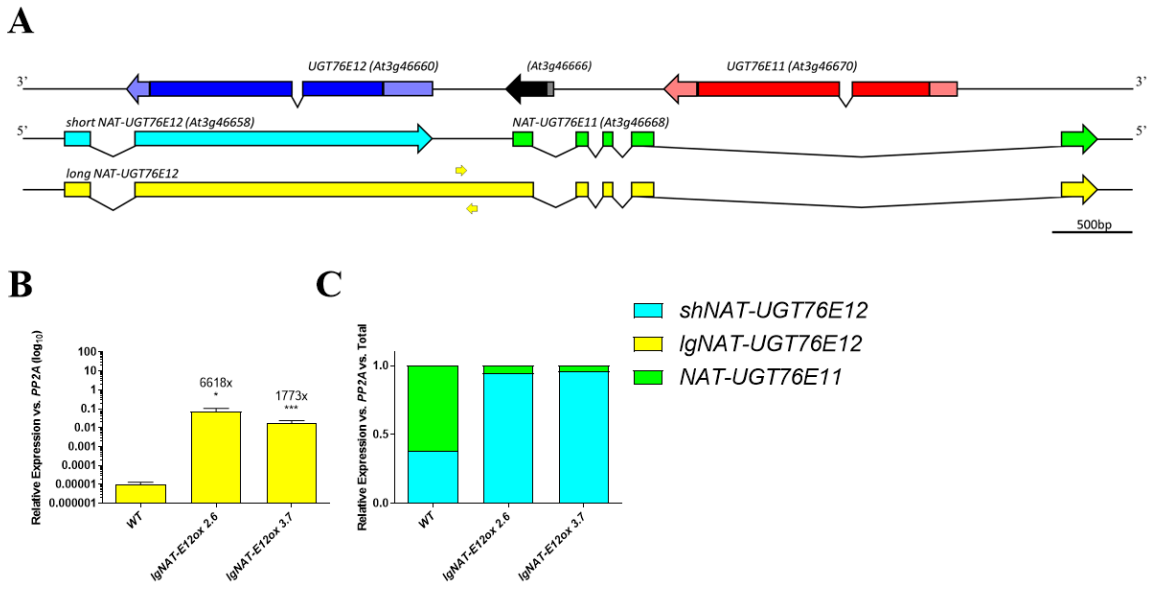


Figure 4.6 Levels of individual NAT in *lgNAT-UGT76E12ox* lines. *A,* Genomic layout showing location of primers used in qRT-PCR (the color of the small arrows corresponds to the target gene). ***B,*** qRT-PCR measuring the expression levels of *lgNAT-UGT76E12* in *lgNAT-UGT76E12ox* lines. ***C,*** Comparison of expression levels of individual NATs. The levels of *lgNAT-UGT76E12* are substantially lower than *shNAT-UGT76E12* and *NAT-UGT76E11* and are not visible. Total RNA extracted from 10 day old seedlings. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. Two-tailed t-test; *p<0.05, **p<0.01, ***p<0.001

In addition to ectopic overexpression lines, we identified a T-DNA line that we called *nat-ugt76e12*^{D} which has increased levels of *NAT-UGT76E12* (**Figure 4.7A, B**). In contrast to ectopic overexpression lines, the T-DNA insertion results in a *cis* upregulation of *NAT-UGT76E12*. When initially analyzing *nat-ugt76e12*^{D}, we observed higher *UGT76E12* expression levels, however, the results between experiments were inconclusive. This is most likely due to the extremely low basal levels of *UGT76E12* which makes measurement difficult. Based on previously published microarray data (Hruz et al., 2008) and our observations, we found *UGT76E12* expression to be upregulated after salt stress. We tested treatments of 250 mM NaCl for 2, 6, and 24 hours and we observed saturation of *UGT76E12* levels after 6 hours with no increases between 6 and 24 hours (**Supplementary Figure 8.10**). Therefore, to facilitate *UGT76E12* quantification in the *nat-ugt76e12* line we induced

expression by treating seedlings for 2 hours to observe differences before saturation could occur. In comparison to *WT* after salt stress, we detected almost twice the level of *NAT-UGT76E12* and *UGT76E12* in *nat-ugt76e12*^{D} (**Figure 4.7B,C**). Since ectopically overexpressed *NAT-UGT76E12* resulted in no differences in sense gene expression, these results suggest that *NAT-UGT76E12* positively regulates *UGT76E12* only in *cis*.

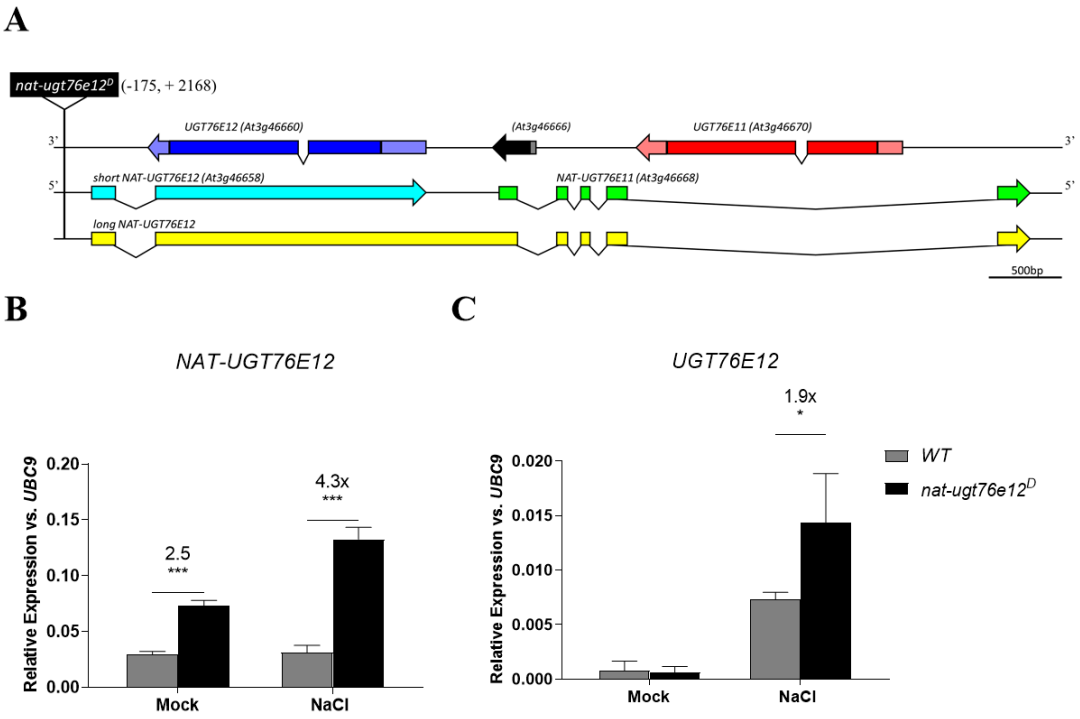


Figure 4.7 Effect of salt stress on expression levels in the *nat-ugt76e12*^{D} T-DNA line. *A*, Insertional mutant line *nat-ugt76e12*^{D} resulting in *NAT-UGT76E12* in upregulation. The position of the insertion is noted in parenthesis. Positive and negative values indicate upstream and downstream of the start codon or to the TIS for protein-coding genes and long non-coding RNA genes, respectively (+1 is the first nucleotide). The insertion is located -175 bp upstream of *NAT-UGT76E12* and + 2168bp downstream of *UGT76E12* ***B,C,*** qRT-PCR of *WT* and *nat-ugt76e12*^{D} 10 day old seedlings grown in liquid media after 2 hours of 250mM NaCl treatment. Mock samples were supplemented with media without salt. Data are mean ± SD from 3-4 biological replicates (n=3-4) from 1 of 2 independent experiments with similar results. Values above bars indicates fold change in comparison to *WT*. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001.

4.5 *Knockout of NAT-UGT76E12 results in decreased UGT76E12 expression*

Previous studies have shown that non-coding RNAs can act in *cis* by having an influence at the gene locus (Chen & Penfield, 2018; Y. Wang et al., 2018). Our inability to observe changes in *UGT76E12* expression upon ectopic overexpression of *NAT-UGT76E12* may suggest a form of *cis* regulation. We utilized the genome editing system CRISPR/Cas9 to generate a knockout of *NAT-UGT76E12* by deleting the first exon and part of the theoretical promoter. We obtained two independent lines *nat-ugt76e12 34.1* and *47.1* with

deletions of 545 (**Figure 4.8A**) and 389 bp (**Supplementary Figure 8.11A**), respectively. Both deletions resulted in a complete knockout of *NAT-UGT76E12* (**Figure 4.8B, C, data not shown for *nat-ugt76e12 47.1***). The levels of *UGT76E12* were drastically reduced indicating that the absence of *NAT-UGT76E12* negatively affects the expression of *UGT76E12* (**Figure 4.8 B,C**) (**Supplementary Figure 8.11B**)**.** This finding, in combination with the results from the *nat-ugt76e12*^{D} line, indicates that *NAT-UGT76E12* could regulate *UGT76E12* expression in *cis*. Furthermore, no changes in *UTG76E11* or *NAT-UGT76E11* were observed in the *nat-ugt76e12* knockout lines indicating their regulation is independent of *NAT-UGT76E12*. In *nat-ugt76e12 34.1*, the deletion is 198 nt downstream of the annotated 3' UTR of *UGT76E12*. While polyadenylation elements are found in the 3' UTR and downstream elements are not present in plants (Xing & Li, 2011), the possibility the 3' end is incorrectly annotated must be excluded. Using 3' RACE, we found the 3' end of *UGT76E12* to be 9 bp shorter than the TAIR annotation and concluded that the deletions do not affect the maturation of *UGT76E12* mRNA (**Supplementary Figure 8.12**).

Additionally, a knockout of *NAT-UGT76E11* was also generated (**Supplementary Figure 8.13A, B**). In this line, neither *UGT76E11* nor *UGT76E12* expression levels were affected and only slightly increase in *NAT-UGT76E12* expression was detected (**Supplementary Figure 8.13C**). Since *UGT76E11* overlaps a spliced intron of *NAT-UTG76E11*, these results further reinforce the requirement of an overlapping region in the transcribed RNA in order regulate gene expression.

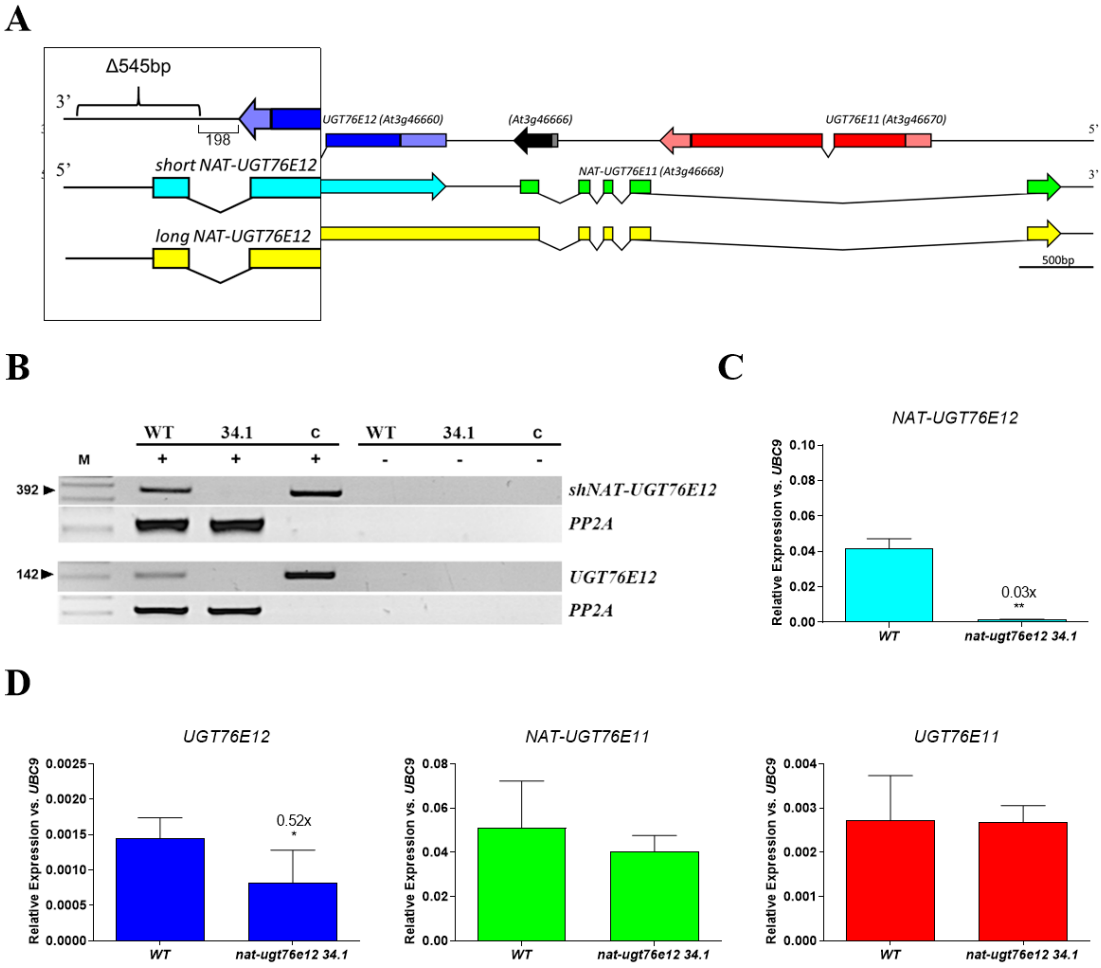


Figure 4.8 Effect of *nat-ugt76e12* knockout on gene expression. ***A,*** Schematic of the *UGT76E12*/*UGT76E11* genomic region showing the 547 bp CRISPR/Cas9 mediated deletion of the first exon and part of the *NAT-UGT76E12* promoter. ***B,*** RT-PCR comparing gene expression in *WT* and *nat-ugt76e12 34.1* knockout line. cDNA reactions with or without reverse transcriptase, +/-; genomic control, c; water control, w. ***C,*** qRT-PCR showing transcript levels in *nat-ugt76e12 34.1* vs. *WT*. ***D,*** qRT-PCR showing the effect of *nat-ugt76e12* knockout on expression of overlapping and nearby genes. Total RNA extracted from 10 day old seedlings and cDNA synthesized using gene specific primers. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001.

4.6 *UGT76E12 RNA stability is unaffected in the absence of NAT-UGT76E12*

Stability and turnover rates are important for maintaining the steady-state RNA levels. These factors greatly vary amongst RNAs and allow rapid changes in gene expression in different developmental and environmental conditions. In order to determine whether *NAT-UGT76E12* influences the stability of *UGT76E12* mRNA we used the transcriptional inhibitor cordycepin. The half-life of *UGT76E12* in *WT* and *nat-ugt76e12* was determined by applying cordycepin and measuring the levels of *UGT76E12* after 40, 80, and 120 minutes. The short-lived *Expansin L1* transcript was included as control (Fedak et al., 2016).

No difference in *UGT76E12* mRNA stability between *WT* and *nat-ugt76e12* was observed further supporting a *cis* regulatory model (**Figure 4.9**).

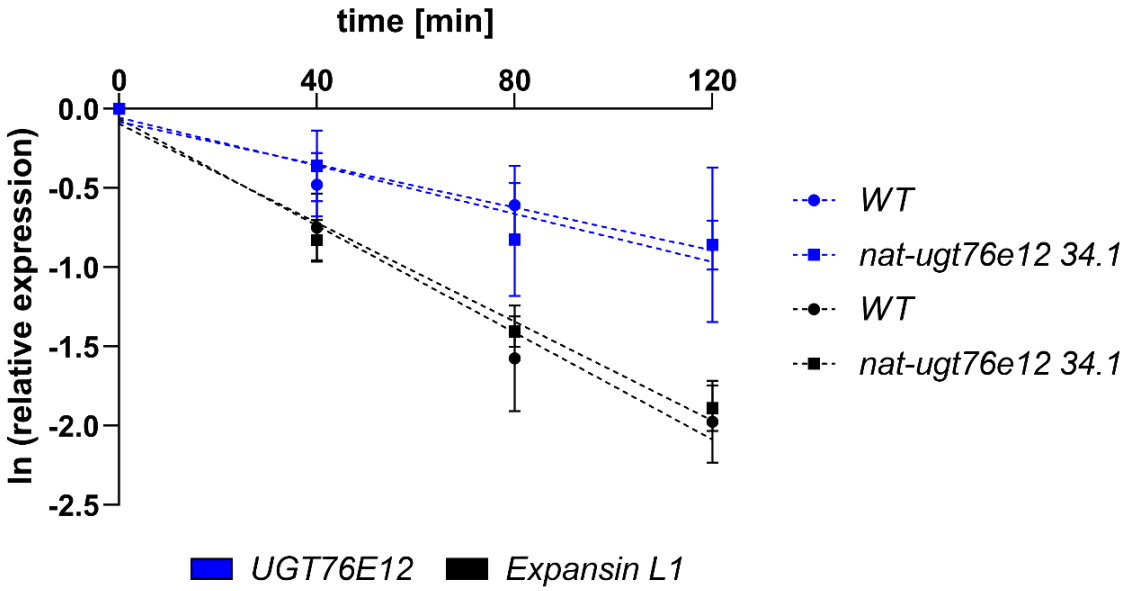


Figure 4.9 RNA stability assay comparing *UGT76E12* levels in *WT* and *nat-ugt76e12* seedlings. The short-lived mRNA *Expansin L1* was used as a positive control for assay functionally. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Dotted line represents best fit.

4.7 *Absence of NAT-UGT76E12 results in altered UGT76E12 DNA methylation*

Previously annotated in a epigenetic database (Lyons & Freeling, 2008), we observed methylated CpG islets throughout the body of *UGT76E12* and the translation initiation site (**Supplementary Figure 8.14A**). Typically in animals DNA methylation is concentrated around the promotor and confers gene suppression, however, patterns in plants are more mosaic-like appearing through the gene body and can also indicate constitutive expression (S. C. Huang & Ecker, 2018; Moore, Le, & Fan, 2013; Sotelo-Silveira, Chavez Montes, Sotelo-Silveira, Marsch-Martinez, & de Folter, 2018). To further investigate the cause of the down regulation of *UGT76E12* in the *nat-ugt76e12* line, we assayed the DNA methylation pattern near the translation initiation site (**Figure 4.10A**). We only observed CpG methylation and in comparison to *WT* plants detected a reduction at several positions closer to the start codon and a slight increase at positions -58 and -65 upstream of the translation initiation site (**Figure 4.10B**). While these results do not directly explain the reduction of *UGT76E12* expression in the *nat-ugt76e12* knockout, it has been shown that DNA methylation is related to the presence of histone modifications and could suggest an altered histone state (Tariq & Paszkowski, 2004). Taken together we conclude that the absence of *NAT-UGT76E12* transcription alters the DNA methylation of *UGT76E12* near the translation initiation site.

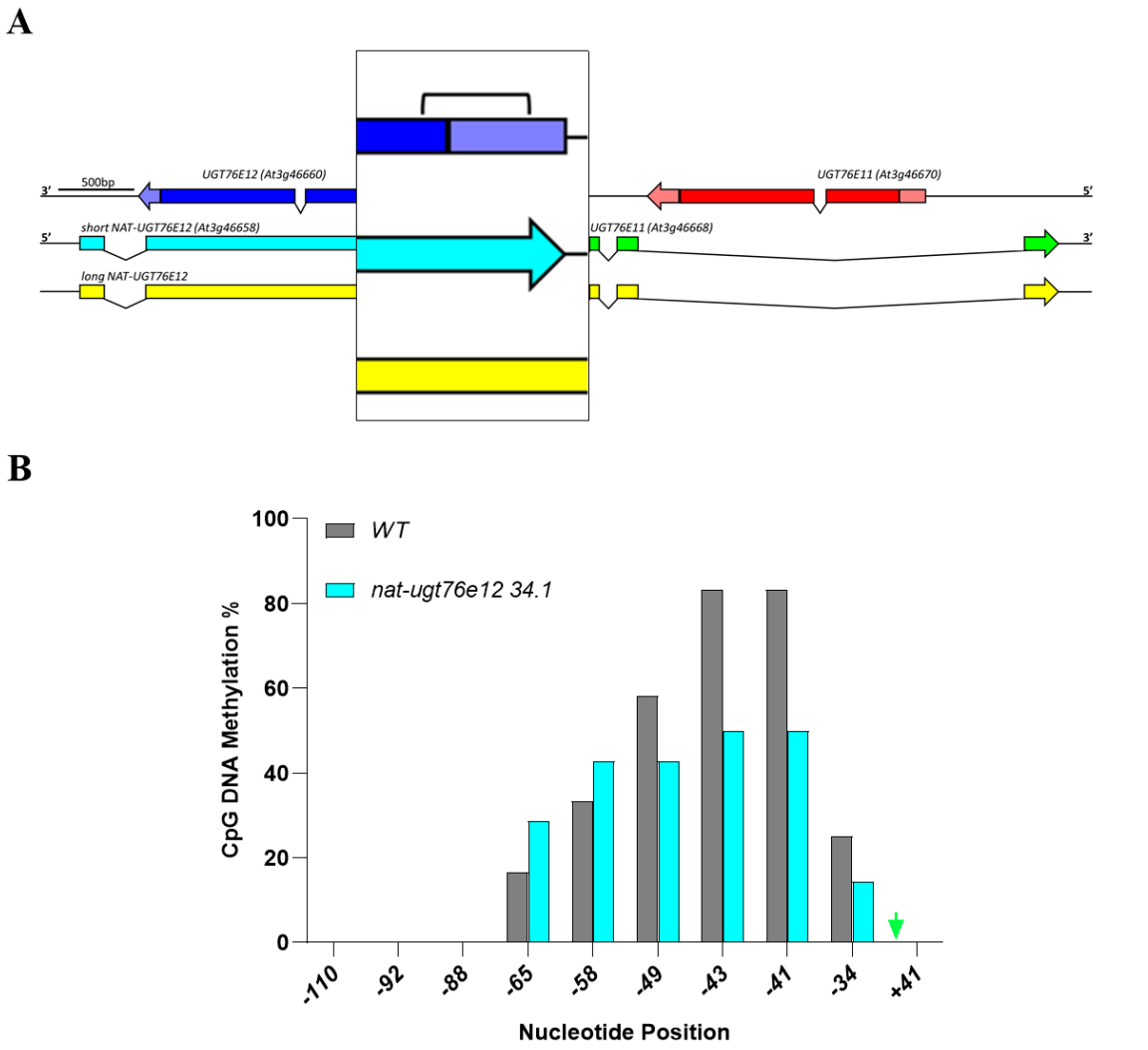


Figure 4.10 DNA methylation surrounding the *UGT76E12* start codon. *A,* Brackets highlight analyzed genomic section: 5'UTR and region surround the start codon. ***B,*** CpG DNA methylation comparing *WT* and *nat-ugt76e12*. The green arrow indicates the start codon (ATG) position. Data are percent of total frequencies of unmethylated and methylated cytosines (n=12-14 sequenced clones). The webtool QUMA was used to quantify DNA methylation (Kumaki, Oda, & Okano, 2008).

4.8 *NAT-UGT76E12 knockout has reduced amount of histone modifications at the 3' end of the UGT76E12 locus*

In recent years, numerous publications have shown in eukaryotes the role long non-coding RNAs have in altering chromatin state, especially histone modifications (Mattick, Amaral, Dinger, Mercer, & Mehler, 2009; Mercer, Dinger, & Mattick, 2009). These mechanisms are poorly understood in plants, however, there are examples of lncNATs influencing the chromatin state of overlapping and neighboring genes (Csorba et al., 2014; Y. Wang et al., 2018). To investigate whether *NAT-UGT76E12* mediates chromatin modifications at the *UGT76E12* locus that may be affecting its transcription, we performed

chromatin immunoprecipitation (ChIP) comparing *WT* and *nat-ugt76e12*. Since the expression of *UGT7E12* is reduced in *nat-ugt76e12* and due to the absence of repressive markers H3K27me1 and me 3 in the ChIP database, we chose to analyze histone markers associated with transcription activation using antibodies against H3K4me3, H3K36me3, H4K5ac, and H3K27ac (**Supplementary Figure 8.14B**) (Roudier et al., 2011). Although previous studies have shown non-coding RNAs can act on neighboring genes (Y. Wang et al., 2018), we focused on the *UGT76E12* locus since only its expression was observed to be affected in the absence of *NAT-UGT76E12*. Primers were designed across three regions of *UGT76E12*: upstream of the TIS (1), 5'UTR (2, 3, 4), and downstream/3'UTR (5, 6) (**Figure 4.11A**). We were able to measure a significant reduction of all activating modifications at the position 6 (3' UTR) and at position 5 (~380 bp upstream of the transcription termination site (TTS)), only levels of H3K4me3, H3K36me3, and H3K27ac were significantly reduced (**Figure 4.11B**). These results provide conclusive evidence that the absence of *NAT-UGT76E12* transcription leads to a reduction of activating histone markers near the 5' end of *NAT-UGT76E12* and reduced *UGT76E12* expression. Therefore, suggesting that active transcription of *NAT-UGT76E12* is required for the full expression of *UGT76E12*. In addition, we included an antibody against H3K27me3, a repressive histone marker, but in accordance with previously reported data (Roudier et al., 2011), we were unable to detect the presence of the modification at the gene locus (data not shown).

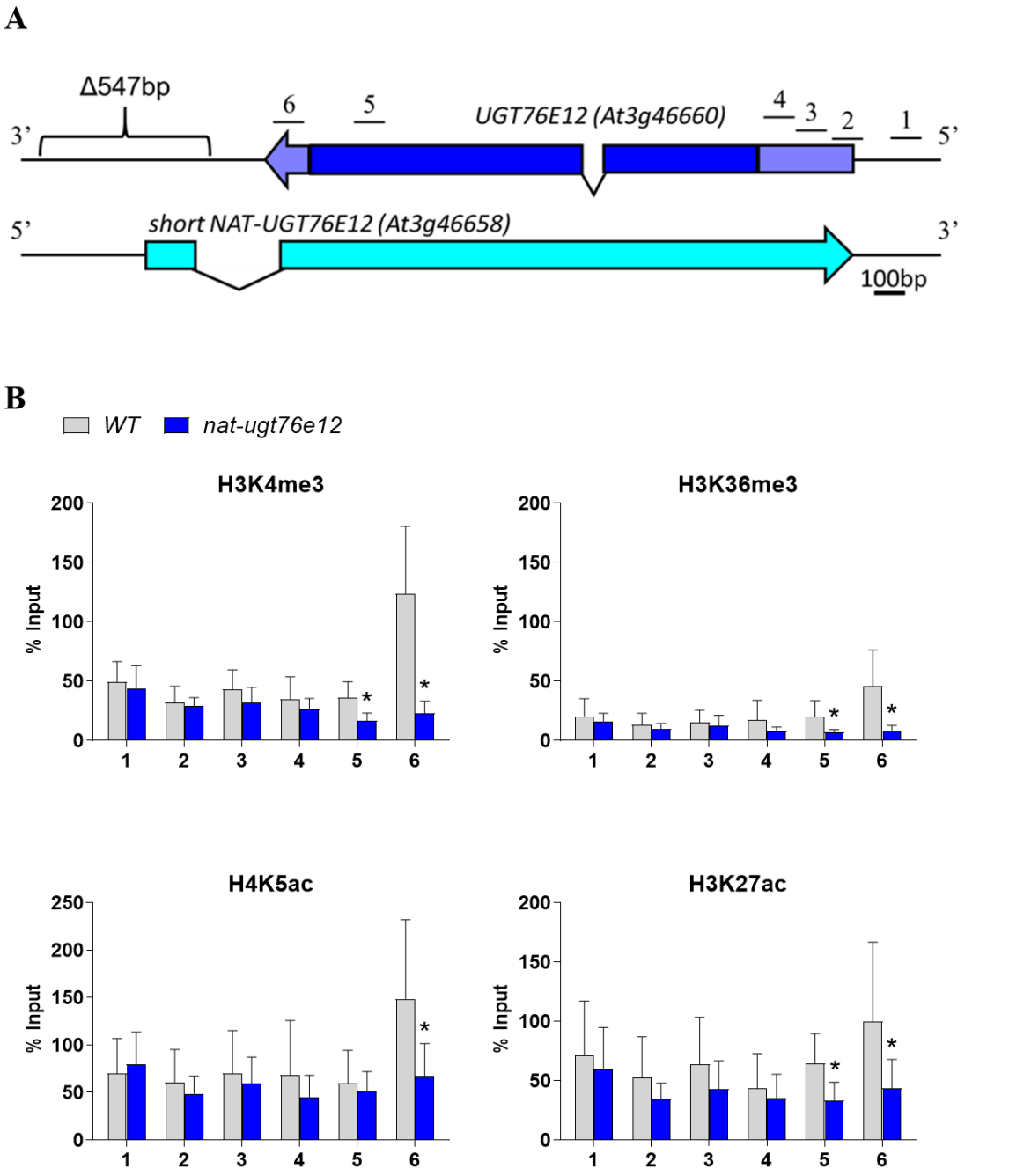


Figure 4.11 Chromatin Immunoprecipitation (ChIP) assay of the *UGT76E12* locus. ***A,*** Schematic representation of the *UGT76E12* locus. The bracket indicates the deleted region in the *nat-ugt76e12 34.1* line. The regions examined are indicated and labeled 1-6. ***B,*** qPCR from ChIP using antibodies against H3K4me3, H3K36me3, H4K5ac, and H3K27ac comparing *WT* and *nat-ugt76e12*. Data are mean ± SD from 2 independent experiments (n=5-7). One-tailed t-test; *p<0.05.

4.9 *UGT76E12 induction is independent of NAT-UGT76E12 expression*

Although *UGT76E12* expression is reduced in *nat-ugt76e12*, the inducibility of *UGT76E12* may be independent of *NAT-UGT76E12* expression. To test whether *UGT76E12* remains inducible in the absence of *NAT-UGT76E12*, we perform a salt stress experiment. As previously shown, in mock conditions the expression of *UGT76E12* is ~50% lower in *nat-ugt76e12* in comparison to *WT*. In 2 independent experiments, after 6 hours of salt stress *UGT76E12* expression was increased in both lines but to a significantly lesser extent in *nat-ugt76e12* (**Figure 4.12**). With these results we concluded that the mechanism of induction of *UGT76E12* is independent of *NAT-UGT76E12*, but that transcription of the NAT is required for full induction.

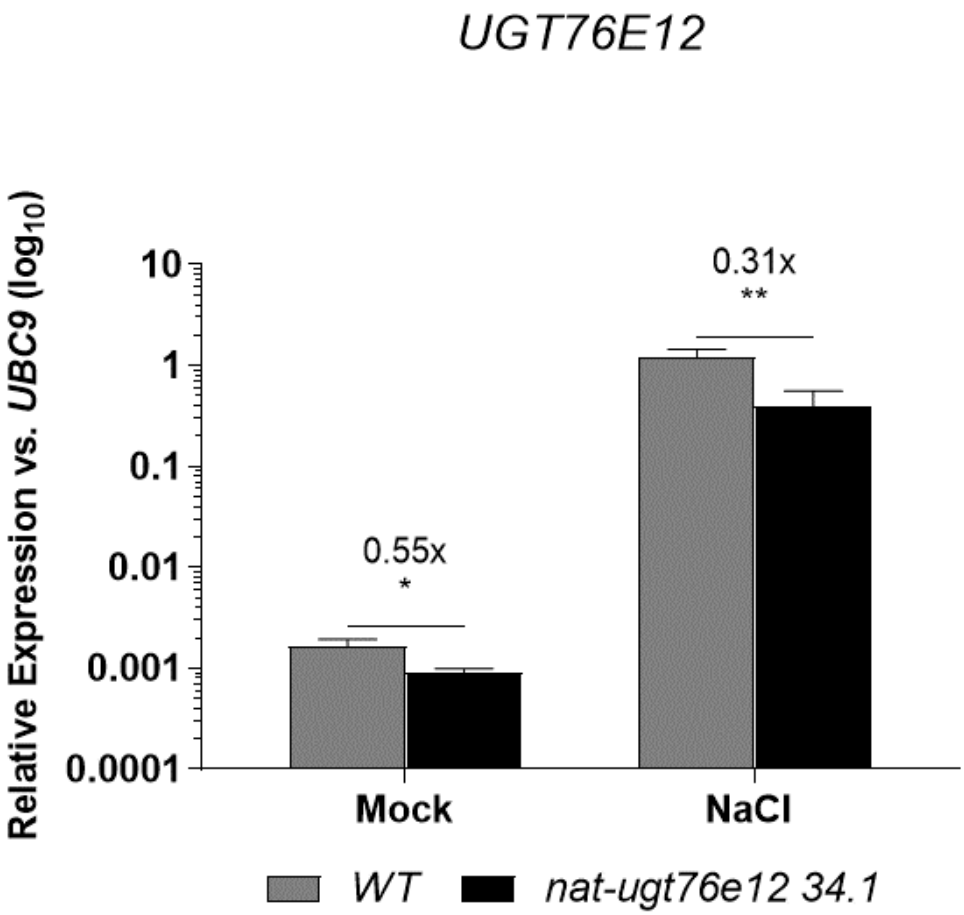


Figure 4.12 Effect of *nat-ugt76e12* knockout on *UGT76E12* expression levels in response to salt stress. qRT-PCR over cDNA from total RNA extracted from *WT* and *nat-ugt76e12 34.1* 10 day old seedlings grown in liquid media after NaCl treatment. Samples were collected after 6 hours of salt stress (250mM NaCl). Mock controls were supplemented with media without salt. Data are mean ± SD from 3-4 biological replicates (n=3-4) from 1 of 2 independent experiments with similar results. Values above bars indicates fold change in comparison to *WT*. Two-tailed t-test; *p<0.05, **p<0.01.

Part II – Analysis of *UGT76E11* and *UGT76E12*

While the main scope of this thesis is to investigate *NAT-UGT76E12*, the function of both protein coding genes *UGT76E11* and *UGT76E12* remains unclear. However, microarray data shows a substantial upregulation upon exposure to biotic and abiotic stresses suggesting a potential role in stress tolerance (Hruz et al., 2008).

4.10 *Salt stress increases the expression of the sense and antisense genes*

We first aimed to validate previously published tiling array data of abiotic stress experiments were Matsui and collaborators showed *UGT76E11* and *UGT76E12* expression increases after a period of exposure to NaCl (Matsui et al., 2010). We were able to confirm this data and observed significant increases in both *UGT76E11* and *UGT76E12* after 6 hours of treatment with 250 mM NaCl of 10 day old seedlings (**Figure 4.13A left**). We additionally measured the expression of *NAT-UGT76E11* and *NAT-UGT76E12* and similarly observed an increase after treatment but to a much lesser extent (**Figure 4.13A right**). We further examined 2 and 24 hours of treatment and found that the expression levels saturate at 6 hours (**Supplementary Figure 8.10**). To observe how salt stress affects promoter activity, we repeated the same NaCl treatment with the reporter gene lines. The promoter of *UGT76E12* showed a clear increase in activity in the shoots and roots. In 10 day old seedlings the promoter of *UGT76E12* is exclusive to roots, however, under stress conditions the domain of activity includes the shoots (**Figure 4.13B**). The low basal expression of *UGT76E12* and its substantial upregulation, surpassing *UGT76E11*, suggests that it may have a more important role in tolerance against salt stress. The absence of visible changes in the *UGT76E11* reporter line in comparison to the fold increase observed qRT-PCR can be a result of the significantly higher sensitivity of the latter that allows the quantification of smaller differences in gene expression. In summary, contrasting the tiling array data by Mastui and collaborators, where they observed decreased *NAT-UGT76E12* levels after salt stress (Matsui et al., 2010), our results show a positive correlation with *UGT76E12*.

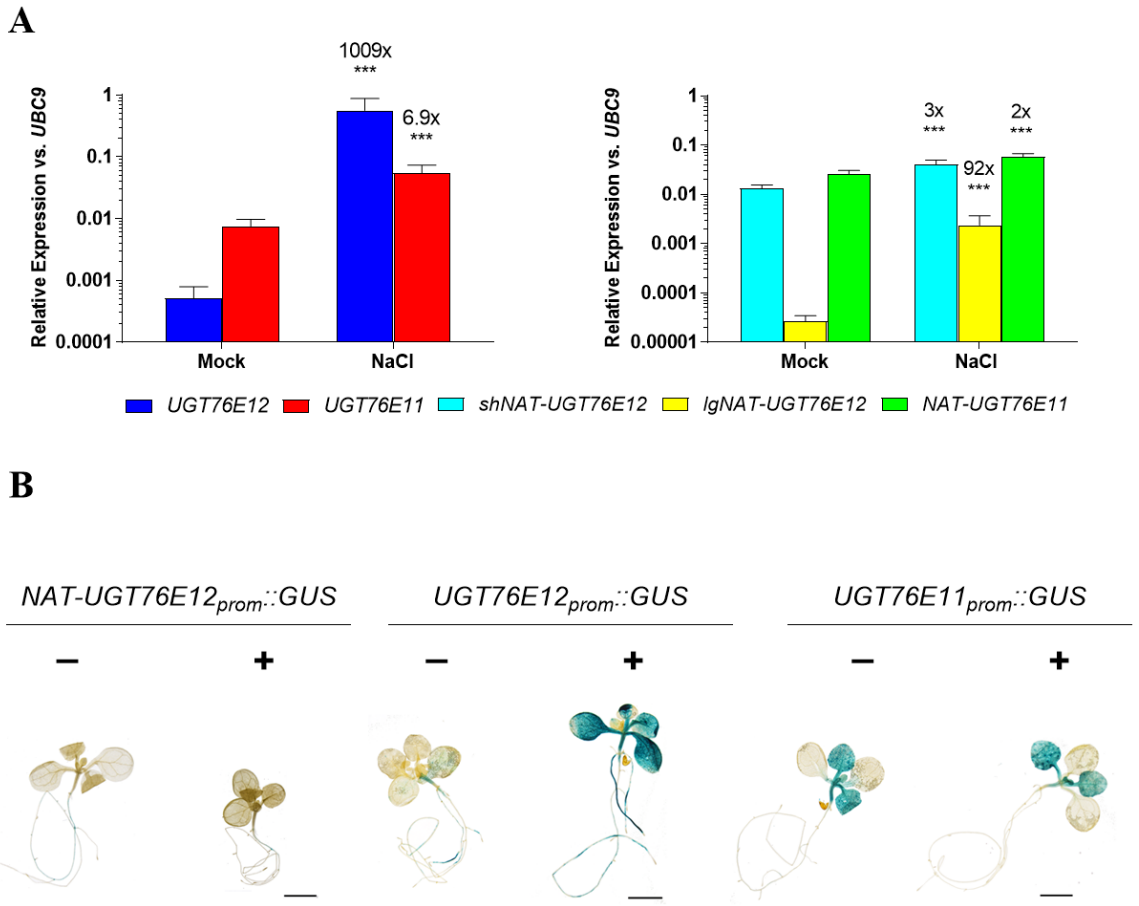


Figure 4.13 Reporter gene activity and expression analysis after salt stress. ***A***, qRT-PCR of Col-0 10 day old seedlings grown in liquid media after 6 hours of 250mM NaCl treatment. Data are mean ± SD from 3 biological replicates (n=3) from 1 of 2 independent experiments with similar results. Values above bars indicates fold change in comparison to mock treatment. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001. ***B***, Histochemical analysis of 10 day old seedling carrying the *promoter::GUS* reporter after 6 hours of 250 mM NaCl treatment (+). (-) indicates the mock treated plants. Mock controls were supplemented with media without salt. Pictures shown correspond to 1 of 3 independent lines. Scale = 1.0 cm

4.11 *Alterations of UGT76E11 and UGT76E12 expression affects susceptibility to B. cinerea infection*

Next, we examined the effect of biotic stress on expression. Based on microarray data, in the 76E sub-family only *UGT76E11* and *UGT7612* are substantially upregulated 48 hours post infection (hpi) with the necrotrophic fungus *B. cinerea* (**Figure 4.14A**). We first aimed to confirm the expression levels in *WT* by infecting 7 day old seedling on plates and collecting samples at 6 and 24 hpi. We observe increases only in *UGT76E12* (**Figure 4.14B left**) and surprisingly decreased *NAT-UGT76E12* expression at both time points and a slight increase in *NAT-UGT76E11* after 24 hpi (**Figure 4.14B right**). The decreased *NAT-UGTE12*

levels had no negative impact on *UGT76E12* expression reinforcing that its induction is independent of *NAT-UGT76E12*. To further examine the potential role of these proteins have in plant defense, we perform infection experiments on overexpression and knockout lines of *UGT76E11* and *UGT76E12*. Both open reading frames (ORFs) were overexpressed by placing them under control of the constitutive 35S cauliflower mosaic virus (CaMV) promoter (**Supplementary Figure 8.15**). We also identified two T-DNA insertion mutant lines, *ugt76e11*^{D} and *ugt76e12*^{D}, which result in an upregulation of *UGT76E11* and *UGT76E12*, respectively (**Supplementary Figure 8.16**). Additionally, double and single knockout lines for *UGT76E11* and *UGT76E12* were created by introducing frameshift mutations using CRISPR/Cas9 (**Supplementary Figure 8.17A**). Since the *ugt76e12* single and *ugt76e11e12* double knock contain a deletion or insertions of 1 to 2 nucleotides at the beginning of the ORF, to ensure any early occurring methionine do not act as start codons and produce truncated protein, an LC/MS/MS analysis of protein extracted from seedlings subjected to salt stress was used to confirm the absence of the protein (**Supplementary Figure 8.17B, C**). During this analysis, the *ugt76e11* knockout was still being generated and not included, however, since the line contained several mutations we assumed that no functional protein could be generated.

Using the overexpression and knockout lines, we infected leaves of 35 day old plants with the pathogen and used the lesion area and the quantity of *B. cinerea* DNA to measure the progression of the infection. Under our conditions no lesions were observed 24hpi and excessive decay at 72hpi made handling of the samples too difficult, consequently, samples were collected 48hpi. Though *UGT76E12* is upregulated after infection in *WT*, suggesting a potential role in resistance, overexpression resulted in larger lesion areas while the lesions in double and single knockout lines were smaller at 48hpi (**Figure 4.15A**).The necrotrophic nature of *B. cinerea* suggests that larger lesions translate to more available nutrients and in turn increased growth. To determine whether the growth of *B. cinerea* was affected and further reinforce the observed phenotype, we quantified the amount of *B. cinerea* DNA in the lesion. In agreement with the lesion area, independent overexpression of both genes resulted in higher quantities of DNA in all of the lines except the low overexpression line *E11ox 4.7* (**Supplementary Figure 8.15 left**) which also displayed a weaker lesion area phenotype (**Figure 4.15B**). In contrast, only the double knockout had a significant reduction of *B. cinerea* DNA. When examining all the data, a positive correlation between *UGT76E12* expression levels and stronger lesion area phenotype was observed in most cases. The lines with significant differences in lesion area but no increased DNA quantity are most likely a

result of the higher chances of error due to the process of collection, extraction, and measurement of the samples. In comparison, lesion area measurement requires minimal manipulation, therefore, reducing the chances of error. Nonetheless, we concluded that both *UGT7E11* and to a higer extent *UGT76E12* play a role in immunity during *B. cinerea* infection and are detrimental when constitutively overexpressed.

We also infected *shNAT-UGT76E12* and *lgNAT-UGT76E12* overexpression lines and the knockout line *nat-ugt76e12*. In comparison to the overexpression of the protein coding genes, the NATs only showed marginally larger lesions but no differences in the quantity of *B. cinerea* DNA. In *nat-ugt76e12*, no differences in lesion area or *B. cinerea* DNA were observed (**Supplementary Figure 8.18A, B**). In combination with the previous sections, these results further support that ectopic overexpression of *NAT-UGT76E12* does not affect the expression of *UGT76E12* and the larger lesions in the *NAT-UGT76E12ox* lines are not a result of altered *UGT76E12* expression. Furthermore, the lack of phenotype in *nat-ugt76e12* is most likely because *UGT76E12* remains inducible (**Figure 4.12**) and its levels are sufficient to avert changes in susceptibility against *B. cinerea*.

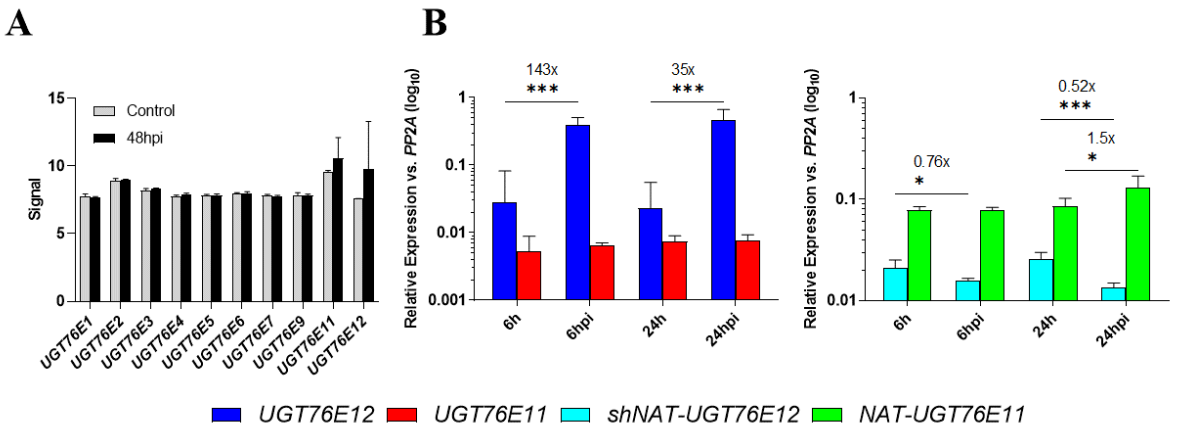


Figure 4.14 Effect of *B. cinerea* infection on gene expression levels. *A,* Histogram of signals of the 76E sub-family members 48hpi with *B. cinerea*. Data from microarray on the Genevestigator database (Hruz et al., 2008). ***B,*** Expression levels determined by qRT-PCR in 7 day old seedlings at 6 and 24 hpi with *B. cinerea*. Seedlings were grown on plates. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to mock treatment. One-tailed t-test; *<0.05, **<0.01, ***<0.001.

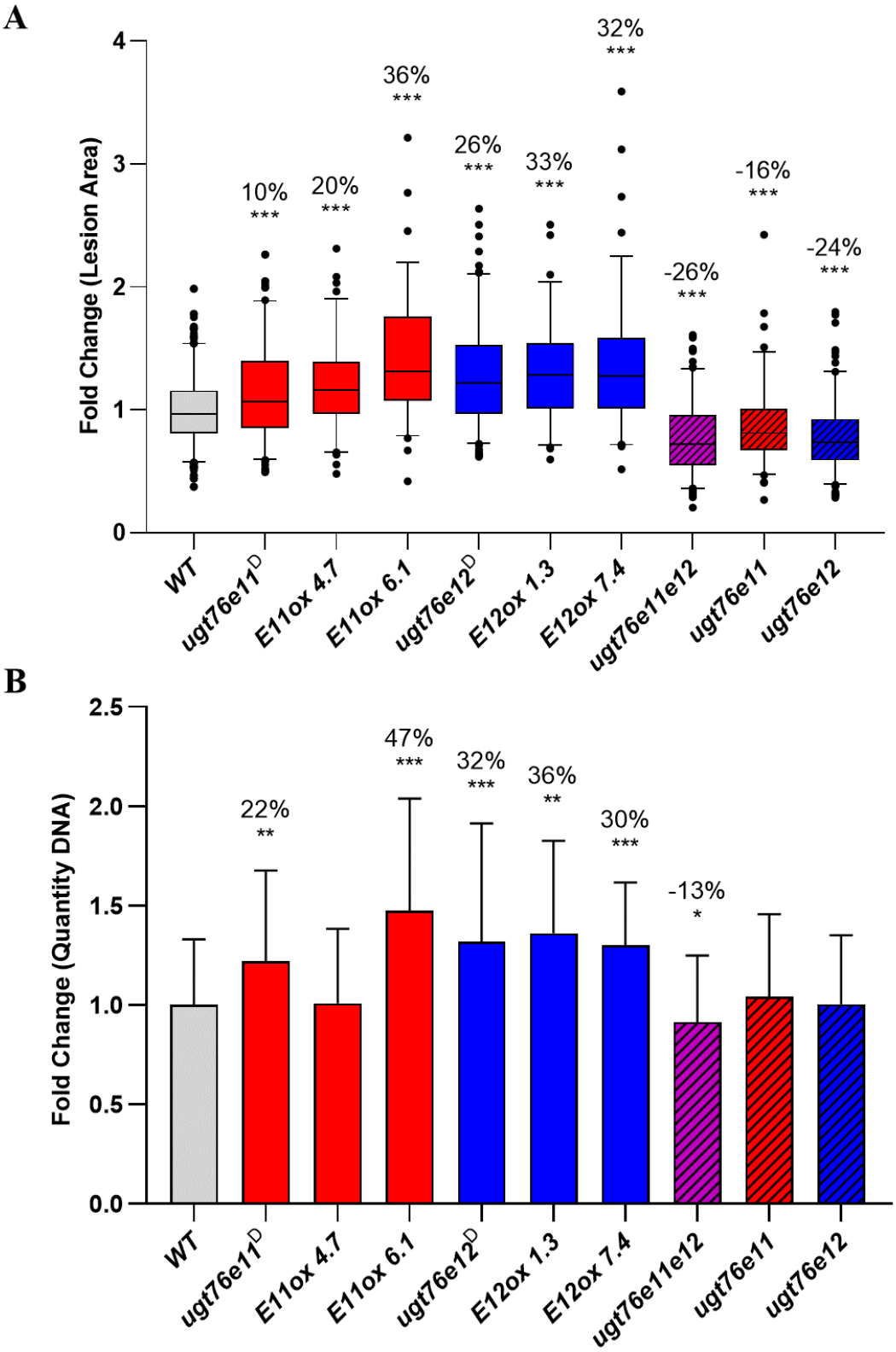


Figure 4.15 The effects of altered *UGT76E11* and *UGT76E12* expression on susceptibility to *B. cinerea* infection. ***A***, Fold change of lesion area measured at 48hpi expressed as percentage compared with the *WT*. Boxplot indicates median and interquartile range. Outliers (0-0.5% and 95-100%) are indicated with black dots. Shown are compiled data from 13 independent experiments (n=72-332). ***B***, Fold change of *B. cinerea* DNA measured using qPCR expressed as percentage compared with the *WT*. Data are mean ± SD from compilation of 13 independent experiments (n=12-58). Values above bars indicates fold change in comparison to *WT*. One-tailed t-test *p<0.05, **p<0.01, ***p<0.001.

5 Discussion

Part I – *NAT-UGT76E12* and *NAT-UGT76E11*

5.1 *Transcript characterization*

In this work we aimed to elucidate the function of the *NAT-UGT76E12* and its potential mechanism of action, however, there exists conflicting data between the TAIR and Arabidopsis Information Portal (Araport) databases regarding the size of the transcript. In order to generate knockout lines and appropriately sized overexpression and reporter gene lines, we needed to analyze the splicing pattern and determine the transcription initiation site (TIS) and transcription termination site (TTS). We consistently observed complete splicing of introns in all three lncNATs at every developmental stage. The TIS for *NAT-UGT76E12* was consistent between the databases and using 5' RACE we observed it to be located 33 nt downstream of the reported site. Our 3'RACE revealed a population of *NAT-UGT76E12* molecules with TTSs located between ~100 to 250 nt downstream of the TAIR10 reported site (**Supplementary Figure 8.1**). This variability is most likely the cause of the discrepancies between the databases, especially when taking into consideration how the annotations were determined. TAIR10 uses expressed sequence tag (EST) sequences and only two RNA-Seq datasets. In comparison, Araport11 combines information from TAIR10 and 113 publicly available tissue-specific RNA-Seq libraries. It also integrates strand-specific RNA-Seq which is critical for accurately annotating NATs (Cheng et al., 2017). In our investigations, we observed a short and long variant of *NAT-UGT76E12* (**Figure 4.1**). The short variant, *shNAT-UGT76E12*, is annotated in TAIR10 and Araport while the long variant, *lgNAT-UGT76E12*, was observed only in a previously published tiling array (Matsui et al., 2010). In this publication, *lgNAT-UGT76E12* was proposed to be 13.8 kB in length and span *UGT76E12*, *UGT76E11*, and 3 other UGTs (**Supplementary Figure 8.2**). However, we observed that *lgNAT-UGT76E12* has a shared TIS with *shNAT-UGT76E12* and a shared TTS with *NAT-UGT76E11*, therefore, only overlapping *UGT76E12* and *UGT76E11* and spanning from *NAT-UGT76E12* to *NAT-UGT76E11*. We performed a 3' RACE on the end shared by *NAT-UGT76E11* and *lgNAT-UGT76E12* and detected difference of only +3 and +5 nts in comparison to the TAIR annotation.

Nonetheless, when this project was started Araport was relatively new and we mostly relied on TAIR10. Although we observed varying 5' and 3' ends from RACE, a portion of the transcripts had the approximate annotated TAIR10 ends (**Figure 4.1**, **Supplementary**

Figure 8.1). Therefore, we decided to overexpress *shNAT-UGT76E12* using the TAIR10 reported ends and in addition *lgNAT-UGT76E12* using the experimentally confirmed ends. Although no changes in the expression of *UGT76E11* and *UGT76E12* were observed in the *shNAT-UGT76E12ox* or *lgNAT-UGT76E12ox* lines, most of the transcripts in the *lgNAT-UGT76E12ox* lines terminated at the *shNAT-UGT76E12 3'* end indicating the presence of a strong transcriptional terminator (**Figure 4.5**, **Figure 4.6**). While *lgNAT-UGT76E12* is expressed in *WT* plants, this may be due to the leakiness of the *shNAT-UGT76E12* TTS and a phenomenon called transcriptional readthrough. Since *NAT-UGT76E11* is located downstream of *shNAT-UGT76E12* and shown to be independently transcribed, some RNA Pol IIs may remain attached after transcription of *shNAT-UGT76E12* and continue until reaching the promoter of *NAT-UGT76E11*. Then RNA Pol II proceeds to transcribe *NAT-UGT76E11* resulting in *lgNAT-UGT76E12*. While not yet studied in plants, in human cells it has been shown that RNA Pol II does not immediately dissociate and can still be attached up to 10 kB downstream of a polyadenylation site (Core, Waterfall, & Lis, 2008). Additionally, Vilborg and collaborators showed that certain stresses to human and mouse cells can lead to an increase in transcriptional readthrough and that this phenomena is dependent on the transcription of the upstream gene and the efficiency of transcription termination (Vilborg, Passarelli, Yario, Tycowski, & Steitz, 2015; Vilborg et al., 2017). The elongated transcript can stay intact and was shown to be retained inside the nucleus and proposed to function as a nuclear scaffold in response to stress (Vilborg & Steitz, 2017). Alternatively, the sequence downstream of the polyadenylation site can be cleaved and degraded due to the unprotected 5' end (Porrua & Libri, 2015). The presence of *lgNAT-UGT76E12* may be a result of transcriptional readthrough but is not subject to degradation or retained in the nucleus due to the TTS and polyadenylation of *NAT-UGT76E11*. We observed that *shNAT-UGT76E12* and *lgNAT-UGT76E12* are upregulated in salt stress and are both absent in the *nat-ugt76e12* knockout lines. This suggests they share the same transcription initiation signals and that the transcription of *lgNAT-UGT76E12* is dependent on transcription of *shNAT-UGT76E12* which may hint to some sort of function. This hypothesis is further reinforced by the precise transcriptional termination of the end shared by *NAT-UGT76E11* and *lgNAT-UGT76E12*. A novel transcript is annotated in Araport11 to be downstream of the TIS of *NAT-UGT76E11*, however, RNA-Seq data shows minimal expression and therefore transcriptional readthrough may not occur. Furthermore, *NAT-UGT76E11* expression is unaffected in the *nat-ugt76e12* knockout line indicating the presence of an independent promoter and differential regulation of the NAT.

Whether *lgNAT-UGT76E12* is biologically relevant or just "transcriptional garbage" remains in question. Nevertheless, RNA-Seq on overexpression lines may reveal independent *trans* functions of *shNAT-UGT76E12* and *lgNAT-UGT76E12*. It could help to detect changes in gene expression which could further assist in elucidating a possible *trans* mechanism. Additionally, one or both transcripts can be involved in interactions with RNA binding proteins or regulate microRNA activity by target mimicry.

5.2 *Promoter activity and expression*

Until recently, NATs in *A. thaliana* have been shown to negatively influence expression of the sense gene (Baurle & Dean, 2006; Fedak et al., 2016; Henriques et al., 2017; Wunderlich et al., 2014). However, a new RNA-Seq study revealed a strong tendency of lncNATs to positively correlate with the expression of their sense gene (X. Zhao et al., 2018). In order to determine how *NAT-UGT76E12* and *UGT76E12* expression relate, we compared their spatiotemporal expression and promotor activities (**Figure 4.3**, **Figure 4.4**, **Supplementary Figure 8.4 - 7.11**). Our time course experiments showed a positive correlation between *NAT-UGT76E12* and *UGT76E12* expression. Furthermore, in stress free conditions *NAT-UGT76E12* transcript levels remained higher than *UGT76E12* at all stages of development and in certain organs. When examining the reporter gene lines at early time points, the promoter of *UGT76E12* in active throughout the seedling and no *NAT-UGT76E12* promoter activity was observed. Only after 10 days we observe promoter activity for *NAT-UGT76E12* in roots and concurrently the domain of promoter activity for *UGT76E12* was shifted to being exclusively in the same organ. These finding suggest that *NAT-UGT76E12* may be involved in controlling the expression of *UGT76E12* only after a specific developmental stage. In our conditions, the first true leaves are seen to be fully emerged at10 days signifying a possible correlation with the change in expression domain. In contradiction with the report gene lines, the expression analysis examining shoot and root showed similar levels of each gene in both organs. This could be caused by the lower sensitivity of the histochemical analysis in comparison to qRT-PCR or due to a high RNA stability. Nonetheless, we concluded that after 10 days both *NAT UGT76E12* and *UGT76E12* have similar a spatiotemporal expression profile.

Additionally, in salt stress both the protein coding genes and NATs are upregulated (**Figure 4.13A**). These results are comparable to what was observed by Zhao and collaborators where they found that the lncNAT *MAF4 ANTISENSE* (*MAS*) and its sense gene *MADS AFFECTING FLOWERING4* (*MAF4*) are upregulated upon exposure to cold.

MAS was shown to be crucial for upregulating *MAF4*, a *FLC* family member involved in the vernalization response (X. Zhao et al., 2018).

This work primarily focused on the gene pair formed by *NAT-UGT76E12* and *UGT76E12* but also simultaneously monitoring *UGT76E11*, a closely related gene, and its antisense long non-coding RNA *NAT-UGT76E11*. Interestingly, in the time course experiment the expression of *NAT-UGT76E11* and *UGT76E11* were found to be negatively correlated. When examining the reporter gene lines and expression data from shoots and roots, we observed *UGT76E11* to be exclusively detected in the shoots. Since the introns of the mature *NAT-UGT76E11* RNA are all spliced and contain no complementary sequence to *UGT76E11*, we assumed that it would not have an influence on *UGT76E11* at the posttranscriptional level. This was further reinforced by results from the analysis of the *nat-ugt76e11* knock line, where no changes in *UGT76E11* expression levels were observed (**Supplementary Figure 8.13**). Based on our observations we hypothesized that changes in *NAT-UGT76E11* expression may be dependent on changes *NAT-UGT76E12* expression. Crisp and collaborators showed in *A. thaliana* that transcription of a gene could lead to an opening of chromatin and result in the transcription of a gene less than 1000 bp downstream. They observed while RNA Pol II can remain attached due to transcriptional readthrough, 88% of the time the intergenic region between the two genes is removed resulting in two correctly sized transcripts (Crisp et al., 2018). Therefore, transcription of *NAT-UGT76E12* and *UGT76E12* may be influencing the chromatin state facilitating the transcription of *NAT-UGT76E11*. Since *NAT-UGT76E11* is only 500 bp downstream of the TTS of *NAT-UGT76E12*, the combination of RNA Pol II readthrough and an active chromatin state might be the cause of the increased *NAT-UGT76E11* transcription levels observed in the time course and salt stress experiments (**Figure 4.3B, Figure 4.13A**). We also observed a slight increase in *NAT-UGT76E11* expression levels in *nat-ugt76e12*^{D} (data not shown) where a T-DNA insertion located -175 bp upstream of the TIS of *NAT-UGT76E12* results in an increased level of *NAT-UGT76E12* expression in *cis*. Taken together, this data suggests that increases in *NAT-UGT76E11* expression levels may be in part dependent on *NAT-UGT76E12* transcription. However, based on the higher expression levels of *NAT-UGT76E11* when compared to *NAT-UGT76E12*, we still concluded that both are independently expressed, each under control of their own promoter.

In addition, using the *nat-ugt76e11* knockout line we performed expression analysis before (25 days) and after (35 days) the intersection point where expression levels of *NAT-UGT76E11* increase and *UGT76E11* levels decrease (**Figure 4.3B**). In a preliminary

experiment we observed no differences in comparison to *WT* plants suggesting that the decrease in *UGT76E11* expression as the plant ages is not regulated by *NAT-UGT76E11* (**Supplementary Figure 8.19**). In future work, promoter reporter gene line for *NAT-UGT76E11* would further confirm the presence of an independent promoter and explain the potential influence of *NAT-UGT76E12* transcription and may facilitate a better understanding of spatiotemporal promoter activity of *NAT-UGT76E11*.

5.3 *Function and potential mechanism of NAT-UGT76E12 action*

In comparison to the other studied lncNATs in plants, the promoter region and 5' end of *NAT-UGT76E12* and *NAT-UGT76E11* do not overlap their sense genes. This characteristic allowed us the advantage of deleting the non-overlapping sequences using CRISPR/Cas9 without affecting *UGT76E11* or *UGT76E12*. The CRISPR/Cas9 deletion lines enabled the investigation of possible *cis* regulatory mechanisms. Moreover, since downstream polyadenylation elements are not reported in plants, the deletion inadvertently affecting transcript maturation was not a concern (Xing & Li, 2011; Z. Zhao, Wu, Ji, Liang, & Li, 2019).

In line *nat-ugt76e12 34.1*, the removal of part of the theoretical promotor and the first exon, resulted in a knockout of *NAT-UGT76E12* expression. In a second line *nat-ugt76e12 47.1*, only the first exon and part of the intron was deleted. In both lines, the absence of *NAT-UGT76E12* resulted in a significant reduction of the basal *UGT76E12* expression levels (**Figure 4.8**, **Supplementary Figure 8.11**). These lines provided the initial evidence that *NAT-UGT76E12* positively regulates *UGT76E12* expression. Additionally, the expression of *NAT-UGT76E11* and *UGT76E11* remained unchanged meaning that *NAT-UGT76E12* specifically regulates *UGT76E12* and not its closely related gene *UGT76E11*. The combination of our observations from *nat-ugt76e12*^{D} where an upregulation in *NAT-UGT76E12* in *cis* results in an increase in *UGT76E12* expression levels (**Figure 4.7**) and the lack of influence of ectopically overexpressed *NAT-UGT76E12* on *UGT76E12* expression (**Figure 4.5**) suggests that *NAT-UGT76E12* mediates *UGT76E12* expression via a *cis* regulatory mechanism. Furthermore, deletion of the first 3 exons of *NAT-UGT76E11* also resulted in a complete knockout of this gene, however, no change in the of expression of *UGT76E11* was observed. Since the mature transcript of *NAT-UGT76E11* contains no complementary sequence to *UGT76E11*, the results suggest that an overlapping region is necessary for NAT regulatory function.

In order to determine the mechanism of how *NAT-UGT76E12* mediates *UGT76E12* expression we primarily utilized the *nat-ugt76e12* knockout line since in this line the expression of *UGT76E12* was significantly affected in both, basal conditions and in response to salt stress (**Figure 4.12**). Because the expression of *UGT76E12* was only reduced in the *nat-ugt76e12* knockout, we first tested whether *NAT-UGT76E12* could have a role in mRNA stability. LncRNAs have been shown to positively and negatively impact mRNA stability. Gong and collaborators showed in human cells that lncRNAs can form duplexes with the 3' UTR of mRNAs and recruit STAUFEN 1 (STAU1), a protein that binds dsRNA, and mediates the degradation of the mRNA (Gong & Maquat, 2011). In contrast, transcripts of *β-SECRETASE-1* (*BACE1*), an important gene in Alzheimer's disease pathophysiology, were shown to be stabilized by its antisense transcript *BACE1-AS* and levels of *BACE1* positively correlated to the expression of *BACE1-AS* (Faghihi et al., 2008). However, the cordycepin based RNA stability assay revealed no changes in *UGT76E12* mRNA half-life in the absence of *NAT-UGT76E12* (**Figure 4.9**). Since no change in *UGT76E12* mRNA levels were observed in *NAT-UGT76E12ox* lines, these results indicate that *NAT-UGT76E12* does not act on *UGT76E12* mRNA stability and points to a *cis* regulatory mechanism modulating transcription.

Since DNA methylation has been shown to influence gene expression and a methylome database showed methylation along the gene body of *UGT76E12* (**Supplementary Figure 8.14**), we hypothesized that *NAT-UGT76E12* may influence the methylation of DNA at the *UGT76E12* locus. In flowering plants, RdDM has been shown to be mediated by transcripts of RNA Pol IV and Pol V, however, Pol II can also play the role of both and independently lead to DNA methylation. In addition, products of RNA Pol II can recruit RNA Pol IV and Pol V and promote RdDM (Matzke, Kanno, & Matzke, 2015; Matzke & Mosher, 2014). Although RdDM is mainly observed in transposon silencing and mostly likely not the mechanism mediating *UGT76E12* methylation, it does provide an example in which non-coding RNAs can change gene expression by modulating chromatin state. We were able to confirm the information from methylome database and detected CpG methylation sites upstream of the *UGT76E12* start codon. In the *nat-ugt76e12* knockout line, we observed a change in the amount of methylation at different positions (**Figure 4.10**). While these results do not directly explain the reason for decreased levels of *UGT76E12*, they indicate a possible change in the chromatin state at the locus. Nonetheless, we observed that the absence of *NAT-UGT76E12*, a product of RNA Pol II, leads to changes in DNA methylation in the 5' UTR near the start codon of *UGT76E12*. However, whether the changes

are a direct or indirect result of the absence of *NAT-UG76E12* transcription remains in question.

Only 3 publications have shown lncRNAs to be involved in histone modifications in plants. Two are lncNATs from *A. thaliana*, *MAS* and *COOLAIR*, and the third is a lincRNA from *Oryza sativa L.*, *LRK ANTISENSE INTERGENIC RNA* (*LAIR*) (Csorba et al., 2014; Y. Wang et al., 2018; X. Zhao et al., 2018). In this work, we describe a third lncNAT of *A. thaliana* that influences the chromatin state. In the *nat-ugt76e12* knockout line, we showed decreases in histone activation markers at and upstream of the 3' UTR of *UGT76E12* (**Figure 4.11**). Like *MAS* and LAIR, *NAT-UGT76E12* positively influences the transcription of the sense gene and exclusively regulates its overlapping/adjacent gene, a common theme amongst lncNAT in plants. While *MAS* is crucial for the cold induced activation of *MAF4* (X. Zhao et al., 2018), *NAT-UGT76E12* is not required for the salt stress induction of *UGT76E12* (**Figure 4.12**). Nevertheless, upon treatment with NaCl *UGT76E12* remained inducible but only to 1/3 of the *WT* levels meaning that *NAT-UGT76E12* is necessary for its full expression.

The mechanism of how *NAT-UGT76E12* influences chromatin state remains to be determined. However, other studied lncRNAs and their involvement in recruitment of histone modifying complexes can give insight into possible mechanisms. While the effect of *COOLAIR* is well understood, the exact mechanism was until recently unknown. Initially, Swiezewski and collaborators observed by fusing the *COOLAIR* promoter to the antisense 3' end of constitutively expressed *GREEN FLOURESENCE PROTEIN* (GFP), that the effects of *COOLAIR* were independent of its sequence and only active transcription was necessary for gene silencing (Swiezewski et al., 2009). They additionally showed that the expression of *COOLAIR* and *FLC* were mutually exclusive at individual loci meaning while one cell can express both, they are always transcribed from different alleles (Rosa et al., 2016). However, a more recent study showed that *COOLAIR* directly interacts with the RNA recognition motif (RRM) of FLOWERING CONTROL LOCUS A (FCA). FCA then interacts with a subunit of the PRC2 complex, CURLY LEAF (CLF), and mediates the trimethylation of H3K27 leading to silencing of *FLC* (Tian et al., 2019). Interestingly, two other lncRNA associated with *FLC*, *COLDAIR* and *COLDWRAP*, were shown to directly associate with PRC2. Unlike *COOLAIR*, *COLDAIR* and *COLDWRAP* are sense transcripts synthesized from an intron and from the promoter of *FLC*, respectively (Heo & Sung, 2011; Kim & Sung, 2017; Kim, Xi, & Sung, 2017). We observed an overall reduction of all histone modifications at position 6 in the *nat-ugt76e12* knockout (**Figure 4.11B**), suggesting *NAT-*

UGT76E12 may be indirectly interacting with modifying complexes. Like *COOLAIR*, *NAT-UGT76E12* may bind to an RBP that associates with the various histone methyl- and acetyltransferases.

Similar to *COOLAIR*, *LAIR* was shown to bind with complexes that acetylate and methylate histones (Y. Wang et al., 2018). Ectopically expressed transcripts of *COOLAIR* and *LAIR* can mediate modifications functioning in a *trans* manner. In contrast, ectopic expression of *MAS* was unable to drive *MAF4* expression (X. Zhao et al., 2018). However, both *LAIR* and *MAS* were reported to bind to a WDR5 protein which forms complexes with histone 3 lysine 4 methyltransferases (Jiang, Kong, Gu, Li, & He, 2011). In addition, both were only able to influence the gene to which their 3' end overlaps. We observed that *NAT-UGT76E12* overexpression in *trans* had no impact on *UGT76E11* or *UGT76E12* expression. While *NAT-UGT76E12* was induced after salt stress, we have no clear evidence correlating the quantity of *NAT-UGT76E12* to the induction of *UGT76E12*. Rather, similar to what was initially reported for *COOLAIR*, the act of transcription may in some way promote the recruitment of modifying complexes but only in *cis*. In addition, presumably due to the splicing of the intron resulting in a lack of a complementary sequence in the mature *NAT-UGT76E11* RNA, no changes in *UGT76E11* were observed in the *nat-ugt76e11* knockout line. While one could argue that the transcription of *NAT-UGT76E11* is simply not involved in the recruitment of histone modifying complexes, a epigenetic database shows similarities in histone modification states between *UGT76E11* and *UGT76E12* (**Supplementary Figure 8.14**) (Roudier et al., 2011). Therefore, further investigation is required to determine the exact mechanism of how histone modifications are mediated at the locus.

Using previously published examples we speculate potential mechanisms in which *NAT-UGT76E12* can function. In the case of *COOLAIR* (Tian et al., 2019) and *LAIR* (Y. Wang et al., 2018), they are able to bind to RBPs in *trans* which then interacts with histone modifying complexes (HMC). *MAS* was also shown to bind to an RBP, but only in *cis* (X. Zhao et al., 2018) (**Figure 5.1A**). In contrast, lncRNAs can also act as scaffolds and are able to directly associate and simultaneously recruit various modifying complexes in *cis* (Joh, Palmieri, Hill, & Motamedi, 2014). Both COLDAIR (Kim et al., 2017) and COLDWRAP (Kim & Sung, 2017) are able to bind to HMC directly and shown to act in *cis* and *trans* (**Figure 5.1B**). Closely resembling our observations, van Werven and collaborators showed in yeast that lncRNA *IRT1* can recruit SET2 and SET3 in *cis* and lead to changes in chromatin state (van Werven et al., 2012). Furthermore, in yeast and mammalian cells the catalytic subunits SET1 and SET2 of H3K4 and H3K36 methyltransferases, respectively,

have been shown to associate directly with RNA Pol II (Tanny, 2014). (**Figure 5.1C**). Since the absence of *NAT-UGT76E12* can influence the level of various histone modification types, it may suggest an indirect associate with histone modifying complexes via other RBP or RNA Pol II. An RNA pulldown of *in vivo* crosslinked samples could be performed to identify potential interactors for *NAT-UGT76E12*. Since DNA methylation and histone modifications can influence each other (Cedar & Bergman, 2009), this would confirm whether *NAT-UGT76E12* may be involved in guiding DNA methylation indirectly altering the histone state or directly mediate histone modifications.

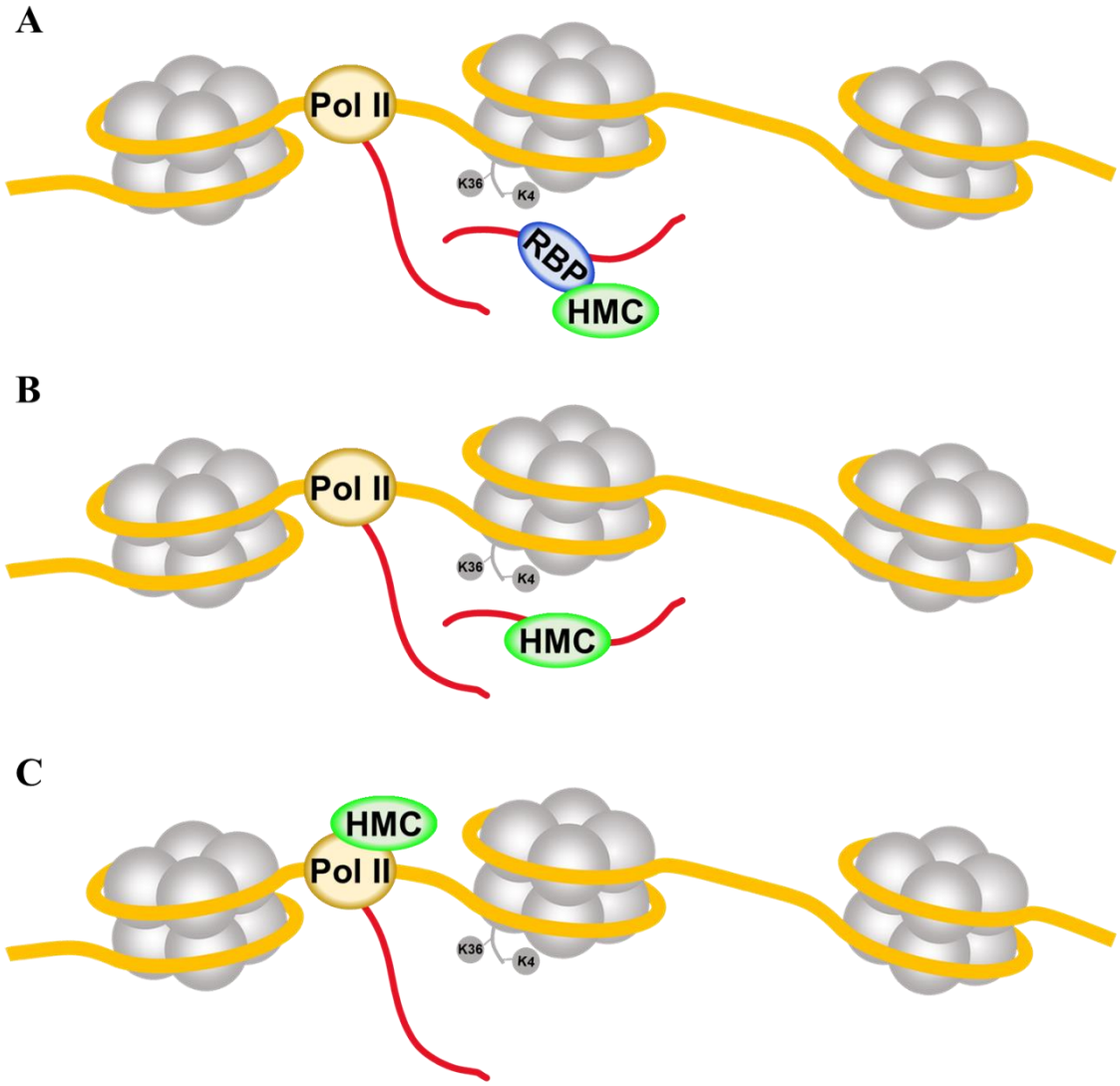


Figure 5.1 Mechanisms of how lncRNAs can recruit HMC. A lncRNA (red) is transcribed by RNA Pol II (yellow) and interacts with ***A***, an RBP (blue) that associates with HMC (green) or ***B***, directly with HMCs. DNA (orange) and histones (grey) ***C***, HMC can also interact with RNA Pol II (Tanny, 2014) and active transcription indirectly recruits them to the locus.

While we cannot be certain that *NAT-UGT76E12* functions in the same fashion as the previously described lncRNAs, we also postulate that active transcription may result in

a mechanical opening of the chromatin and indirectly affect the chromatin state. This hypothesis stems from the ChIP assay where in the *nat-ugt76e12* knockout line we observed a reduction in histone markers but only toward the 3' end of *UGT76E12* which is near the TIS of *NAT-UGT76E12* (**Figure 4.11**). In contrast, all previously mentioned examples showed changes near the 5' ends of the sense gene suggesting their mode of action may differ from that of *NAT-UGT76E12*. In summary, while the precise mechanism of *NAT-UGT7612* remains to be determined, we conclude that its absence results in a reduction of histone activation markers at the 3' end and in an altered CpG DNA methylation near the start codon of *UGT76E12*.

Part II – *UGT76E12* and *UGT76E11*

Until recently only limited information was known about the function of *UGT76E11* and *UGT76E12*. Using our experimental findings along with publicly available microarray databases and newly published data, we aimed to elucidate the biological role of these enzymes.

5.4 *UGT76E subfamily*

The 76E12 subfamily, part of group H, contains 10 members located on either chromosome 3 (*UGT76E3*, *UGT76E4*, *UGT76E5*, *UGT76E6*, *UGT76E11*, *UGT76E12*). or 5 (*UGT76E1*, *UGT76E2*, *UGT76E7*, *UGT76E9*). While many still require investigation, some have been characterized *in vitro* and/or *in vivo*. Using *in vitro* assays UGT76E1, UGT76E2, UGT76E11, and UGT76E12 have been observed to glycosylate quercetin and two oxylipins (11-hydroxy7,9,13-hexadecatrienoic acid (11-HHT) and 13- hydroxy-9,11,15-octadecatrienoic acid (13-HOT)) (Haroth et al., 2019; Lim et al., 2004). The only with clear enzymatic activity *in vivo* is *UGT76E1*, which accumulates after wounding and glycosylates 12-hydroxy-jasmonic acid (12-*O*-JA) yielding 12-*O*-glucopyranosyl-jasmonic acid. While *UGT76E2* and *UGT76E12* also displayed activity toward 12-*O*-JA *in vitro, in vivo* activity was not observed (Haroth et al., 2019). Nevertheless, potential *in vivo* substrates for UGT76E11 and UGT76E12 still need to be identified. However, a recent publication used a chemical-bioinformatic model and predicted candidate substrates that could be experimentally tested (Yang et al., 2018).

5.5 *Functions of UGT76E12* and *UGT76E11*

Tiling- and micro array data showed that *UGT76E11* and *UGT76E12* are upregulated after salt stress and *B. cinerea* infection (Hruz et al., 2008). We found that *UGT76E12*,

having a low basal level, was significantly more upregulated in comparison to *UGT76E11* after 6h of salt stress (**Figure 4.13A**). Under salt stress, Li and collaborators were able to observe an enhanced germination, increased root length, and increased survival rate in *UGT76E11* overexpression lines (Q. Li et al., 2018). However, experiments performed in our laboratory revealed that *UGT76E12* may have a larger role in germination under salt stress conditions (Gentile, personal communication). Our results indicate that the single knockout of *ugt76e12* and double knockout of *ugt76e11e12* but not the single knockout of *ugt76e11* had lower germination rates under salt stress (Gentile, personal communication). Furthermore, between the T-DNA lines *ugt76e11*^{D} and *ugt76e12*^{D} only the line with enhanced *UGT76E12* expression had a higher rate of germination in salt stress conditions (Gentile, personal communication). The root length was also affected but only the double knockout showed significantly reduced root length which also showed a significantly reduced green area in comparison to *WT* in drought stress experiments (Gentile, personal communication). These results suggested that *UGT76E12* and *UGT76E11* may share a common/similar substrate and in some cases able to compensate one another.

In order to elucidate the role of *UGT76E12* and *UGT76E11* in *B. cinerea* infection, we performed infection experiments using overexpression and knockout lines. Surprisingly, overexpression of both genes resulted in increased susceptibility to infection and single and double knockout lines showed decreased susceptibility (**Figure 4.15**). In both cases, alterations in *UGT76E12* expression had a more profound effect in comparison to changes in *UGT76E11*. All the *UGT76E12ox* lines and the *ugt76e12*^{D} line had larger lesions and higher amounts of *B. cinerea* DNA. Moreover, in the *ugt76e12* knockout we observed smaller lesions similar to the double *ugt76e11e12* knockout but DNA levels were significantly decreased only in the double knockout. While altered *UGT76E11* levels have similar consequences, the effects were to a much lesser extent compared to *UGT76E12.* When considering the post-infection expression levels, *UGT76E12* seems to have a more crucial role as it is the only one that is upregulated (**Figure 4.14B**). These results contradict our initial hypothesis that since *UGT76E12* is upregulated after infection, its overexpression would provide increased protection against *B. cinerea*. However, the low basal levels of *UGT76E12* in *WT* plants could suggest that high expression levels are detrimental and potentially leads to decreased leaf integrity resulting in an increased susceptibility to infection in the presence of the pathogen. *B. cinerea* can also manipulate the expression of host genes by releasing compounds that antagonize necrotrophic fungal defense pathways (El Oirdi et al., 2011) and siRNAs that can silence defense related genes (Weiberg et al.,

2013). *UGT76E12* may be upregulated by the fungus in order to weaken plant defenses by an unknown mechanism.

While *NAT-UGT76E12ox* lines showed no altered *UGT76E11* or *UGT76E12* expression, we performed infection experiments to assay if they potentially have an effect in *trans* that might be related to necrotrophic fungal defense. We observed only marginal increases in lesion area for most lines, but no changes in the quantity of *B. cinerea* DNA (**Supplementary Figure 8.18**). Overall, differences in DNA quantity were mostly seen in lines with larger differences in lesion area. The reduced increase in lesion area and lack of differences in DNA quantity in these lines could mean that *NAT-UGT76E12* may have a related but less influential function in *trans.* We also infected the *nat-ugt76e12* knockout and observed no difference in lesion area or *B. cinerea* DNA quantity (**Supplementary Figure 8.18**). While the *UGT76E12* levels are significantly reduced in the *nat-ugt76e12* knockout it remains inducible, but does not reach the levels detect in *WT* plants (**Figure 4.12**). Even so, the level of *UGT76E12* in *nat-ugt76e12* after *B. cinerea* infection is enough to see a phenotype similar to *WT* plants.

The observed phenotypes may be caused by a disruption of the JA pathway based on *in vitro* assays showing UGT76E12 activity towards 12-*O*-JA (Haroth et al., 2019) but the activity still needs to be shown *in vivo*. Von Saint Paul and collaborators showed altered *UGT76B1* expression to have a strong effect on transcription levels of salicylic acid (SA) and jasmonic acid (JA) related genes (von Saint Paul et al., 2011). They observed that a knockout of *UGT76B1* lead to increased expression of SA related genes and a decrease in JA related genes resulting in increased resistance to biotrophic pathogens but also to an increased susceptibility toward nectrophic pathogens. In contrast, overexpression of *UGT76B1* resulted in the opposite effect suggesting it may directly influence both SA and JA pathways (von Saint Paul et al., 2011). However, our preliminary data from lines with altered expression of *UGT76E12* and *UGT76E11* showed no clear effect on the expression levels of genes of both pathways (data not shown). Future experiments using samples from salt stress and *B. cinerea* infections in targeted and non-targeted metabolite analysis along with *in silico* predictions may help in identifying the substrates of UGT76E12 and UGT76E11 *in vivo*.

In summary, we showed that both *UGT76E11* and *UGT76E12* play a role in *B. cinerea* infection and salt stress tolerance. Our results suggest that mainly *UGT76E12* may be involved in the cross talk between biotic and abiotic stresses. The basal expression levels of *UGT76E11* are higher in comparison to *UGT76E12* and decreases as the plant ages

suggesting it may be more related to developmental processes (**Figure 4.3**). *UGT76E12* may play a role in linking the abiotic abscisic acid (ABA) mediated pathway involved in salt stress to the biotic JA mediated pathway involved in necrotrophic pathogen infection. Performing infections using the biotroph *Pseudomonas syringae* will help further elucidate whether *UGT76E12* may also be involved in SA signaling. Nonetheless, the interactions and crosstalk between biotic and abiotic defense signaling pathways is wildly complex and requires in depth study.

6 Summary and conclusions

Our investigation of *NAT-UGT76E12* gives a first example of a lncNAT in *A. thaliana* positively influencing a sense gene involved in stress response. Although, no phenotype was seen in *B. cinerea* infection experiments using the *nat-ugt76e12* knockout (**Supplementary Figure 8.18**), a clear decrease in *UGT76E12* levels was observed in the absence of *NAT-UGT76E12* (**Figure 4.8C**, **Supplementary Figure 8.11B**). Since the inducibility of *UGT76E12* remains unaffected in the *nat-ugt76e12* knockout and in the T-DNA line *ugt76e12*^{D} (where the insertion is -8 bp upstream of the ATG of *UGT76E12*) (**Supplementary Figure 8.20**), the upregulation in stress conditions may possibly be due to the binding of an activating transcription factor or to the release of a repressor at a site at or near the ATG. A schematic summarizing our results is shown in **Figure 6.1**.

(1) Based on RACE, the 3' end of *shNAT-UGT76E12* can be up to 245 nt longer than the TAIR10 annotation and result in various sized transcripts. The 5' annotated end was consistently 33 nt shorter than the TAIR10 annotation.
(2) The long variant *lgNAT-UGT76E12* has lower levels of expression in comparison to *shNAT-UGT76E12* and is most likely produced due to transcriptional readthrough.
(3) *NAT-UGT76E11* is independently expressed and under the control of its own promoter. *NAT-UGT76E11* overlaps *UGT76E11* across an intronic region and does not influence its expression levels.
(4) *NAT-UGT76E12* positively affects *UGT76E12* expression only in *cis*. The effect is exclusive to *UGT76E12* and does not influence the expression of the closely related and neighboring gene *UGT76E11*.

(5) Infection with the necrotroph *B. cinerea* increases *UGT76E12* and decreases *NAT-UGT76E12* expression. Overexpression of both *UGT76E12* and *UGT76E11* increases susceptibility to *B.cinerea* infection and the double knockout of these genes had increased resistance.

(6) Salt stress increases expression levels of both sense and antisense genes. Relative to the other genes, *UGT76E12* expression is highly induced. In contrast to necrotrophic pathogen infection, double knockout of *UGT76E12* and *UGT76E11* has reduced tolerance to salt stress and lines with increased expression of *UGT76E12* showed enhanced tolerance.

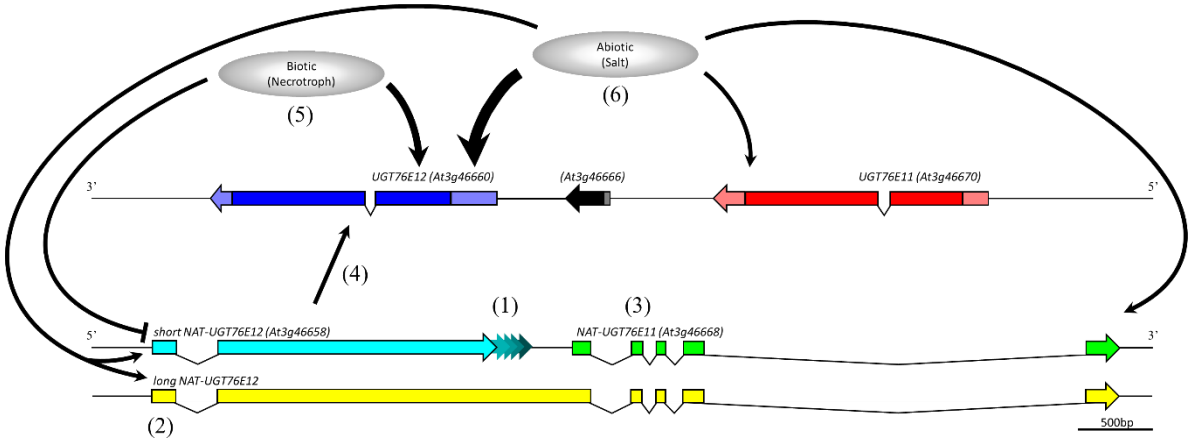


Figure 6.1 Proposed mechanism of lnNAT's action and gene responses to stress. Thickness of the arrow implies the approximate strength of response. A thicker arrow indicates a stronger effect.

Our results advocate that *NAT-UGT76E12* influences *UGT76E12* transcription only in *cis* and does not affect the expression of the closely related gene *UGT76E11*. In addition, the mechanism involved in the induction of *UGT76E12* is independent of *NAT-UGT76E12*. However, *NAT-UGT76E12* is required to reach both, basal *UGT76E12* expression levels and full-induction in response to salt stress conditions. The effect of *NAT-UGT76E12* on *UGT76E12* expression is modulated by changes in DNA methylation at the *UGT76E12* 5' end and by histone modifications at the 3' end.

7 Materials and Methods

7.1 *Gateway cloning and plasmid vectors*

All the transgenic lines in this work were generated using the Gateway Cloning system (Thermo Fisher Scientific) with the plasmids list in **Table 7.1**. Unless stated otherwise, the entry clones were generated using pENTR™ Directional TOPO cloning Kit (Thermo Fisher Scientific) and the expression clones were made using the Gateway™ LR Clonase Enzyme Mix (Thermo Fisher Scientific) and the selected destination vectors according the manufacturer's instructions. Plasmids were isolated using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific) and transformed into *Escherichia coli* TOP10 chemo competent (Thermo Fisher Scientific) or *Agrobacterium tumefaciens* GV3101 electro competent cells.

Table 7.1 Plasmids for transgenic lines

Plasmid	Features	References
pENTR/SD/D-TOPO	• *att1* and *att2* recombination sites • Kanamycin resistance	Thermo Fisher Scientific
pB7WG2 (destination vector)	• 35S:: • Spectinomycin resistance • Basta® resistance	(Karimi, De Meyer, & Hilson, 2005; Karimi, Inze, & Depicker, 2002)
pB7WGY2 (destination vector)	• 35S::N-terminal YFP • Spectinomycin resistance • Basta® resistance	
pBGWFS7 (destination vector)	• Promoter::GUS • Spectinomycin resistance • Basta® resistance	

7.2 *Plant lines*

The model plant *A. thaliana* ecotype *Col-0* was used for all experiments in this work. The overexpression, reporter gene, and CRISPR knockout lines were generated in the Col-0 background. T-DNA lines are also in the Col-0 and were obtain from Nottingham Arabidopsis Stock Centre (NASC).

7.3 *In vitro growth conditions*

A. thaliana seeds were surface sterilized for 10 minutes by immersion in 1.3% NaClO and 0.05% Triton X100, followed by 3 washes with sterile distilled water. Seeds were sown on 1/2 Murashige and Skoog (MS) medium containing 1.5% (w/v) sucrose at a pH of 5.6 with 0.8% (w/v) or 1.5% (w/v) phytoagar. After 2-3 days stratification in the dark at 4ºC,

plates or Erlenmeyer flasks containing seeds were transferred to growth chambers set to 22ºC with a light intensity between 85-100 μmol $m^{-2}s^{-2}$, at either long- (16h light / 8h dark) or half-day (12h light / 12hr dark) conditions. Seeds grown directly on soil were treated in a similar manner excluding sterilization.

7.4 *Liquid culture growth conditions*

A. thaliana seeds were surface sterilized and added to 300 mL wide-mouthed Erlenmeyer flasks containing 100mL of sterile 1/2 MS medium and 1.5% (w/v) sucrose at a pH of 5.6. After 2 days stratification in the dark at 4ºC, flasks were transferred to a growth chamber set to 22ºC with a light intensity between 85-100 μmol $m^{-2}s^{-2}$ at long-day conditions (16h light / 8hr dark) and with circular agitation at 85rpm.

7.5 *A. thaliana transformation*

A modified floral dip method was used to stably transform *A. thaliana* (Logemann, Birkenbihl, Ulker, & Somssich, 2006). *A. tumefaciens* transformed with an expression vector of interest was grown on LB medium plates containing the corresponding antibiotics. After two days of incubation at 28°C, the *Agrobacterium* was scraped off and re-suspended in liquid LB and diluted to an OD_{600} of 2.0. A solution of 5.0 % (w/v) sucrose was prepared using tap water and was added to the cell suspension in a 4:1 ratio. Finally, Silwet-L77 was added to a final concentration of 0.03 % (v/v). Budding flowers of ~25-35 days old plants were immersed into the cell suspension and gently agitated for 15 seconds to ensure complete saturation. Plants were then laid horizontality on trays and kept moist by wrapping the tray with plastic wrap overnight.

7.6 *Selection of transgenic plants*

The overexpression and reporter gene vectors also contained the *bar* gene which conveys glufosinate resistance (Basta®) allowing for easy selection. T1 seeds were harvested from transformed plants and densely sown on soil, stratified, and grown in the greenhouse. Circa 5-7 days after germination, seedlings were sprayed with 80 mg/L Basta® solution every two to three days until Basta®-resistant seedlings were visible. The resistant seedlings (T1) were transferred to individual pots and cultivated further in the greenhouse for seed collection. Approximately ten independent T1 lines per construct were isolated, and T1 lines were numbered (#1-#10). Considering that *Agrobacterium*-mediated transformation procedure may lead to multiple insertions of the transgene into the plant genome (Gelvin, 2003), approximately a hundred T2 seeds were sown on ½MS media plates containing 10

µg/mL Basta® and transformants showing a 3:1 segregation in the T2 generation were selected to avoid the presence of more than one insertion. Roughly 10-12 T2 seedlings from each line were further propagated until collection of T3 seeds. As above, T3 seeds were sown on Basta® plates to select for homozygous lines (all seedlings resistant to Basta®). The T3 and T4 seeds were then used for all subsequent experiments.

7.7 *RNA isolation*

Unless stated otherwise RNA isolation was performed using column purification with the Plant RNA Kit, peqGOLD (Peqlab). RNA concentrations were measured using TECAN Infinite® 200 spectrophotometer and the NanoQuant Plate™. Absorbance measurements were made at 260 nm and 280 nm to determine RNA concentration and purity. Extractions with a 260/280 ratio of approximately 2.0 were considered contaminant-free and used for further experiments.

7.8 *Complementary DNA synthesis (cDNA)*

Complementary DNA (cDNA) was synthesized using extracted RNA meant for target transcript quantification by qRT-PCR or detection by RT-PCR. Each cDNA reaction contained 2 µg total RNA and cDNA synthesis was performed using SuperScript II Reverse Transcriptase (Thermo Fisher Scientific). A modified version of manufacturer's instructions was used. In order to remove any residual DNA contamination, 4 µg of total RNA were treated with dsDNase (Thermo Fisher Scientific) in 10 µL or 20 µL reactions according the manufacturer's specifications. Subsequent steps were carried out in one reaction and then split at the end before the addition of the reverse transcriptase (RT). The cDNA reaction used either 500 ng oligo(dT)$_{12\text{-}18}$ or 2 pmole gene-specific primer (GSP) (**Table 7.2**). Reverse primer for the *PP2AA3* reference gene was additionally included when the cDNA reactions were performed with GSP. RNaseOUT^{TM} was substituted with RiboLock RNase Inhibitor (Thermo Fisher Scientific). Each RNA sample results in two cDNA reactions one with and one without reverse transcriptase (RT). The final reaction volume of 20 µL has a total of 2 µg RNA and since 1 µL (200 units) of RT is enough for 5 µg total RNA, most of the reactions required only 0.5 µL of RT.

Table 7.2 Gene specific primers for cDNA synthesis.

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR415	GACAAAAGTCTTCATAGAGTCCTC
UGT76E12 (AT3G46660)	RR413	ACAACTGTCTTCATAGAGTCCTT
shNAT-UGT76E12 (AT3G46658)	RR430	CTCGGACCAATACAGTTTCTG
PP2AA3 (AT1G13320)	RR434	CAGGACCAAACTCTTCAGCAAGACGC

7.9 *Polymerase Chain Reaction (PCR)*

PCRs were performed using Phusion™ High Fidelity DNA Polymerase (Thermo Fisher Scientific) in 20 µL reactions according to the manufacture's specifications. The primers used are listed in **Table 7.3** and 1 µL of a 1/10 dilution of cDNA was used as template. The cycle annealing temperature was calculated by adding 3ºC to the Tm of the primer with the lower anneal temperature and the extension time was determined based on a rate of 40 sec/kB. The reactions were then analyzed using agarose gel electrophoresis.

Table 7.3 Primers for RT-PCR

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR414(F)	CACCATGGAGGAAAAGCCGGCGGGCAG
	RR415(R)	GACAAAAGTCTTCATAGAGTCCTC
UGT76E12 (AT3G46660)	RR430(F)	CTCGGACCAATACAGTTTCTG
	RR508(R)	GTTTGTTATTCTGCAGGTGAG
NAT-UGT76E11 (AT3G46668)	RR388(F)	CACTAACCAGACCCACCTGCG
	RR390(R)	CGAGGCTTACCACTAACTCC
shNAT-UGT76E12 (AT3G46658)	RR499(F)	CACCAAGATCCCTCCTTTATCGGTTTTC
	RR463(R)	GGAGTTCCAATGATCTGCAGG
lgNAT-UGT76E12	RR500(F)	CACCGTATTGGTTTTTTTCTCTTCGTCTTCC
	RR503(R)	GAACATTGATTTTTATTGACCCTAC
PP2AA3 (AT1G13320)	RR433(F)	GCTGTAGGACCGGAGCCAACTAG
	RR434(R)	CAGGACCAAACTCTTCAGCAAGACGC
RT-PCR Primers for CRISPR lines		
UGT76E12 (AT3G46660)	RR711(F)	CTTTCGCTTGCCGCTCTGTATTTG
	RR712(R)	GCAAACCGTGAAACTGGAAAGTC
shNAT-UGT76E12 (AT3G46658)	RR508(F)	GTTTGTTATTCTGCAGGTGAG
	RR463(R)	GGAGTTCCAATGATCTGCAGG

7.10 *Quantitative Real Time PCR (qRT-PCR)*

Fast SYBR Green Mastermix (Thermo Fisher Scientific) was used for qRT-PCRs according to the manufacturer's specifications. The reactions were adjusted to 10 µL and 1 µL of direct undiluted cDNA was used as template. The PCRs were run on the QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific) on the default setting for the Fast SYBR Green Mastermix. The primers are listed in **Table 7.4** and were designed to have an annealing temperature of ~60ºC. Exon-exon junction primers were ordered to be

PAGE purified to ensure the primers were of appropriate length. The efficiency of the primers was determined using a standard curve. The Ct values and the log of each sample dilution was used to determine the slope and the efficiency was calculation with the following equation:

$$Primer\ Efficiency = 10^{\frac{-1}{slope}}$$

A value of 2 indicates 100% efficiency.

7.10.1 Data analysis

The mean Ct value of 2 to 3 technical replicates was used for quantification. The target genes were quantified relative to the reference genes *PP2AA3* or *UBC9*. The normalized relative expression (ΔCt) and the fold change (ΔΔCt) were graphed.

$$Normalized\ \Delta Ct = Primer\ Efficiency^{-(Ct_{target} - Ct_{Reference})}$$

$$Normalized\ \Delta\Delta Ct = Primer\ Efficiency^{-(\Delta Ct_{target} - \Delta Ct_{Reference})}$$

Table 7.4 Primers for qRT-PCR.

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR491(F)	CCAAGTTGTTTTTAGAGAAAGTATCATATC
	RR492(R)	GTTAGCGAACACAATGAAATGGC
UGT76E12 (AT3G46660)	RR766(F)	GTCCAAGCTCCCTTGAAAGAAAC
	RR767(R)	CTAATGATGCAAACCGTGAAACTGG
NAT-UGT76E11 (AT3G46668)	RR435(F)	GAGTGTAGTATGAGCCGCAAAG
	RR436(R)	GAGGCACAAGCTTCCCAGTTC
shNAT-UGT76E12 (AT3G46658)	RR499(F)	CACCAAGATCCCTCCTTTATCGGTTTTC
	RR393(R)	CTCACCTCTTTTATCTCTTTCGGAGTC
lgNAT-UGT76E12	RR511(F)	CATGTTGCGTTTTGTCCCTAAC
	RR431(R)	CAGAATCGGTATAATGGAAATGTC
PP2AA3 (AT1G13320)	DM195(F)	AGCCAACTAGGACGGATCTGGT
	DM196(R)	GCTATCCGAACTTCTGCCTCATTA
UBC9 (AT4G27960)	RR639(F)	CATGTACAAGACAGACAAGAACAAG
	RR640(R)	CTTCCTTAAGGACAGTATTTGTGTCAG
qRT-PCR Primers for CRISPR lines		
NAT-UGT76E11 (AT3G46668)	RR462(F)	CGGGGAAGATGGTTGATTTGGC
	RR436(R)	GAGGCACAAGCTTCCCAGTTC
shNAT-UGT76E12 (AT3G46658)	RR463(F)	GAAAGCTCTCTTCCTCATCTCC
	RR474(R)	GGAGTTCCAATGATCTGCAGG

7.11 *Rapid amplification of cDNA ends (RACE)*

Total RNA was extracted from wild-type seedlings via column purification and cDNA was synthesized using 5 µg of RNA. The cDNA synthesis was followed as mentioned in section 7.8 except using an anchored oligo(dT) primer for the 3'RACE and GSP primers for the 5'RACE. The anchored oligo(dT) contains two added regions for the nested PCR, the first underlined and the second highlighted in green (**Table 7.5**).

7.11.1 3'RACE

Two PCR reactions were performed using the primers listed in **Table 7.5**. The first PCR used cDNA as template and a subsequence nested PCR was performed using the first PCR product as template. The reactions were then analyzed using agarose and polyacrylamide gel electrophoresis. Any visible bands were extracted using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and cloned into the pZErO-2 vector (Thermo Fisher Scientific) using T4 DNA ligase (Thermo Fisher Scientific) according to the manufacturer's specifications. The ligated vectors were then transformed into chemo competent TOP10 cells and plated on LB media with the appropriate antibiotic for selection (Kanamycin). Colony PCR using DreamTaq DNA polymerase (Thermo Fisher Scientific) was performed with M13 forward (-21) and reverse (-29) primers according to the manufacturer's specifications. Positive clones were grown in liquid media for plasmid isolation and sent to Eurofins Genomics for sequencing.

7.11.2 Circular 5'RACE

cDNA primers were first phosphorylated using T4 Polynucleotide Kinase (Thermo Fisher Scientific). The reaction was incubated at for 30 min at 37°C then diluted by 1/10 to obtain a final concentration of 5pmol/µL.

Primer Phosphorylation reaction

- Primer [100 pmol/µL] — 12.5 µL [1250 pmol]
- Polynucleotide Kinase 10X Buffer A — 2.5 µL
- ATP [10 mM] — 1 µL [0.4 mM]
- T4 Polynucleotide Kinase [10 U/µL] — 0.5 µL
- H_20 — util 25 µL

The RT reaction uses SuperScriptIII to allow for higher reaction temperatures to increase primer binding specificity. A total of 5 µg RNA and 5 pmol phosphorylated primer was incubated for 1.5 min at 98°C then transferred to 55°C to anneal the primers. Then the following components were added to the reaction tube and the mixture incubated for 45 min at 55°C:

RT reaction

- 5X RT First-Strand Buffer 2 µL
- DTT [0.1 M] 0.5 µL
- dNTPs [10 mM each] 0.5 µL
- RiboLock Rnase Inhibitor [40 U/µL] 0.5 µL
- SuperScriptIII [40 U/µL] 0.5 µL

To stop the reaction the temperature was increased to 60°C and the mixture incubated for 10 min followed by incubations at 65°C and 70°C for 5 min and 15 min, respectively. Lastly, to remove RNA from the RNA/DNA hybrids generated during the cDNA reaction, 1µL of RNA H [1U/ µL] was added and incubated for 60 min at 37°C. After digestion only single stranded DNA (ssDNA) with a phosphorylated 5'end remain which were subsequently circularized using T4 RNA ligase [Thermo Fisher Scientific].

Circularization/Ligation reaction

- 10X T4 RNA ligase Buffer 10 µL
- T4 RNA ligase [10 U/µL] 5 µL
- H_20 until 100 µL

The reaction was incubated overnight at 30°C. The following day the circularized ssDNA was precipitated using Sodium Chloride-Tris-EDTA buffer (STE).

STE precipitation

- Ligation reaction 100 µL
- Absolute Ethanol 200 µL
- 10X STE buffer 36 µL
- H_20 23.2 µL
- Gylcoblue 0.5 µL

The mixture was incubated overnight at -20°C then centrifuged for 30 min at maximum speed at 4°C. The pellet was washed once with 400 µL of 70% absolute ethanol, centrifuged again for 30 min at 4°C, then dried for 10 min at room temperature, and finally resuspend with 20 µL RNase free water.

Nested PCR

The nested PCR and cloning were performed in the same way as for the 3'RACE mentioned in section **7.11.1**. The primers used for the PCR are listed in **Table 7.5**.

Table 7.5 Primers for Rapid amplification of cDNA ends (RACE)

cDNA synthesis		
Target	**Primer**	**Sequence (5'→3')**
3'RACE	RR82	*GCTCGCGAGCGCGTTTAAAC*GCGCACGCGT TTTTTTTTTTTTTTTTVN
UGT76E11 (AT3G46670)	RR415	GACAAAAGTCTTCATAGAGTCCTC
UGT76E12 (AT3G46660)	RR413	ACAACTGTCTTCATAGAGTCCTT
shNAT-UGT76E12 (AT3G46658)	RR430	CTCGGACCAATACAGTTTCTG
NAT-UGT76E12 (5' RACE)	RR463	GGAGTTCCAATGATCTGCAGG
3'RACE PCR		
Target	**Primer**	**Sequence (5'→3')**
1st anchor sequence	RR83	*GCTCGCGAGCGCGTTTAAAC*
2nd anchor sequence	RR84	GCGTTTAAACGCGCACGCG
UGT76E12 (AT3G46660)	RR633(F1)	GTTGTGACGAAGAGAGCCAAC
	RR430(F2)	CTCGGACCAATACAGTTTCTG
shNAT-UGT76E12 (AT3G46658)	RR382(F1)	GGCTCTCTTCGTCACAAC
	RR383(F2)	GGTTCTCTTGCCGCATTTCGC
lgNAT-UGT76E12	RR457(F1)	GAACTGGGAAGCTTGTGCCTCTAC
	RR398(F2)	GGAGTTAGTGGTAAGCCTCG
5'RACE PCR		
NAT-UGT76E12	RR474(F1)	GAAAGCTCTCTTCCTCATCTCC
	RR475(R1)	GGAGTCTTTGACTGAGATTCTCTCC
	RR476(F2)	CGTCAACCATTAACCTCTTCACAGC
	RR477(R2)	GCTACGACGATACAAAACCACC

7.12 *Histochemical staining*

The expression vector pBGWFS7 (**Table 7.1**) was used to generate transgenic *A. thaliana* plants reporting promotor activity. This plasmid contains the *β-glucuronidase* (*GUS*) gene, which encodes a protein that converts its colorless substrate, X-Gluc, to a visible blue product. The GUS gene is placed under the control of the theoretical promoter of a gene allowing spatiotemporal monitoring of promoter activity. The primers used to generate the constructs are listed in **Table 7.6**. In the constructs for the protein coding genes, *UGT76E11* and *UGT76E12*, approximately 1 kB upstream of the start codon was used. In *NAT-UGT76E12* construct, 1 kB and 2 kB upstream of the transcription initiation site was used. *A. thaliana* was transformed and propagated to T3 homozygous lines as mention in the previous sections. Independent lines showing the same pattern of promoter activity were chosen and 3 were selected for subsequent analysis.

Histochemical GUS staining was carried out with the *Promoter::GUS* transgenic lines. Plant material was incubated in the staining solution overnight at 37ºC.

GUS staining solution

- NaH_2PO_4 (pH 7) 100 mM
- EDTA (pH 7) 10 mM
- Ferricyanid $Fe(CN)_6^{3-}$ 0.5 mM
- Ferrocyanid $Fe(CN)_6^{4-}$ 0.5 mM
- Triton-X 100 0.1%
- X-Gluc 1 mg/mL
- H_20 to final volume

The staining solution was decanted, and the chlorophyll was removed using ethanol. Solutions with increasing concentrations of ethanol (20%, 35%, 50%, 70%) were added and samples incubated at each concentration for 1 h at room temperature and then left overnight in 70% ethanol at 4ºC. If required, additional washes with 70% ethanol were performed until all the chlorophyll had been removed. Samples were the visualized on a Nikon SMZ1270 optical microscope and analyzed with NS Elements software.

Table 7.6 Primers for reporter lines.

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR453(F)	CACCCAGTTCTTAAAAGAACATAATGAAC
	RR454(R)	TTCGAAAATTTACAAGAACTGCTCTGC
UGT76E12 (AT3G46660)	RR455(F)	CACCACTTCTCAGACTCCCGGTTC
	RR456(R)	GAACTACTAATTAAAAGTATCCTGCTCAGC
NAT-UGT76E12-1kb (AT3G46658)	RR484(F)	CACCGCTTAATCTCTCTCTCTTTCTC
	RR485(R)	GGAGGAAGACGAAGAGAAAAAAACC
NAT-UGT76E12-2kb (AT3G46658)	RR483(F)	CACCTGAAACCCTATTCTCCAAAC
	RR485(R)	GGAGGAAGACGAAGAGAAAAAAACC

7.13 *Overexpression constructs*

The expression vector pB7WG2 (**Table 7.1**) was used to generate transgenic *A. thaliana* plants overexpressing the genes of interest. The most notable feature of the vector is the presence of the strong 35S promoter of cauliflower mosaic virus (CaMV) under which the expression of the gene of interest is controlled. The genomic sequence, introns included, of the non-coding RNAs, *shNAT-UGT76E12* and *lgNAT-UGT76E12*, were independently overexpressed. In addition, plants independently overexpressing the coding sequence (CDS) of the protein coding genes were created. The primers used to generate the constructs are listed in **Table 7.7**. *A. thaliana* was transformed and propagated to T3 homozygous lines as mention in sections **7.5** and **7.6**. To quantify the level of overexpression, RNA extraction and cDNA synthesis followed by qRT-PCR was performed as mentioned in sections **7.7**,

7.8, and **7.10**. Independent lines with varying levels of overexpression were chosen and 2 to 5 were selected for subsequent analysis.

Table 7.7 Primers for overexpression lines.

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR414(F)	CACCATGGAGGAAAAGCCGGCGGGCAG
	RR415(R)	GACAAAAGTCTTCATAGAGTCCTC
UGT76E12 (AT3G46660)	RR412(F)	CACCATGCAGGTTTTGGGAATGGAGG
	RR413(R)	ACAACTGTCTTCATAGAGTCCTT
shNAT-UGT76E12 (AT3G46658)	RR500(F)	CACCGTATTGGTTTTTTTCTCTTCGTCTTCC
	RR347(R)	CCTTTGGCACGTTTCTCA
lgNAT-UGT76E12	RR500(F)	CACCGTATTGGTTTTTTTCTCTTCGTCTTCC
	RR503(R)	GAACATTGATTTTTATTGACCCTAC

7.14 *CRISPR/Cas9 Editing*

The vectors used were obtained from Dr. Mily Ron (University California, Davis) (

Table *7.8*) and guide RNAs (gRNA) (**Table 7.9**) were generated using CHOPCHOP (http://chopchop.cbu.uib.no/). A first PCR was performed with Phusion™ High Fidelity DNA Polymerase (Thermo Fisher Scientific) using the primers from **Table 7.9** and the scaffold plasmid as a template. The assembly of the pENTR construct was performed via digestion/ligation reaction using BbsI (Thermo Fisher Scientific) and T4 DNA ligase (Fermentas), the AtU6gRNA pENTR plasmid (

Table *7.8*), and the purified PCR products. Reaction products were transformed into chemo competent TOP10 *E. coli*. Plasmids were extracted from obtained clones and the identity of the inserts and the absence mutations confirmed by sequencing. The LR reaction was performed with Clonase II (Thermo Fisher Scientific) as specified by the manufacture into pMR333, a vector containing *RPS5A::Cas9* (*Cas9* under the control of a constitutively expressed ribosomal promoter) and *OLE1::GFP* (*GFP* under the control of a seed coat promoter for transformants selection). The vector was then agro-transformed into *A. thaliana*, T1 seeds expressing GFP in the seed coat were selected and plants screened for the presence of a deletion via PCR genotyping. T2 seeds not expressing GFP in the seed coat were selected and seedlings genotyped to confirm homozygosity and the absence of the Cas9 gene. Plant DNA was extracted using a solution containing 200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, and 0.5% SDS. The genotyping PCR was performed using DreamTaq (Thermo Fisher Scientific) and the primers used are listed in

Table 7.10. All mutations were confirmed by sequencing the PCR products at Eurofins Genomics.

Table 7.8 Plasmids for CRISPR/Cas9.

Plasmid	Function	Reference
Scaffold	PCR Template	Dr. Mily Ron (unpublished)
AtU6gRNA	pENTR	
pMR333 (CRISPR/Cas9)	pDEST	

Table 7.9 Primers for CRISPR/Cas9 construct assembly.

Construct	Target	gRNA	Primer	Sequence (5'→3')
#1	*UGT76E11* (AT3G46670)	GAAATGGAGGAAAAGCCGGC	RR590(F1)	CACCGAAGACCTATTGAACAAAGCACCAGTGGTC
			RR598(R1)	ATATGAAGACCGCTTCTCCTTGCATGCACCAGCCGGGAATCG
	UGT76E12 (AT3G46660)	TGCAAGGAGAAGCGTAGTGT	RR599(F2)	TATAGAAGACGAGAAGCGTAGTGTGTTTTAGAGCTAGAAATA
			RR600(R2)	ATATGAAGACAGAAACGCCGGCTTTTCCTCCATTTCTGCACCAGCCGGGAATCG
#2	*NAT-UGT76E12* (AT3G46658)	TAGAGAAAATCATCGAATTT	RR590(F1)	CACCGAAGACCTATTGAACAAAGCACCAGTGGTC
			RR604(R1)	ATATGAAGACGATGATTTTCTCTATGCACCAGCCGGGAATCG
		GACTCCGAAAGAGATAAAAG	RR605(F2)	TATAGAAGACAAATCATCGAATTTGTTTTAGAGCTAGAAATA
			RR603(R2)	ATATGAAGACAGAAACCTTTTATCTCTTTCGGAGTCTGCACCAGCCGGGAATCG
#3	*NAT-UGT76E11* (AT3G46668)	TGAACCGGGAGTCTGAGAAG	RR590(F1)	CACCGAAGACCTATTGAACAAAGCACCAGTGGTC
			RR737(R1)	ATATGAAGACAGACTCCCGGTTCATGCACCAGCCGGGAATCG
		TTCCTCACTATCTTCTGCAA	RR738(F2)	TATAGAAGACGGGAGTCTGAGAAGGTTTTAGAGCTAGAAATA
			RR739(R2)	ATATGAAGACAGAAACTTGCAGAAGATAGTGAGGAATGCACCAGCCGGGAATCG

Table 7.10 Primers for CRISPR/Cas9 genotyping

Gene	Primer	Sequence (5'→3')
UGT76E11 (AT3G46670)	RR631(F)	GATCGGTCACAGGGAAGAAAAG
	RR632(R)	CAGTGAAGTCATCTGAAGGGC
UGT76E12 (AT3G46660)	RR633(F)	GTTGTGACGAAGAGAGCCAAC
	RR634(R)	GTATTGGTCCGAGATTCTTG
NAT-UGT76E12 (AT3G46658)	RR484(F)	CACCGCTTAATCTCTCTCTCTTTCTC
	RR407(R)	CTAAAACACACAAAGCATCAAGATTCC
NAT-UGT76E11 (AT3G46668)	RR511(F)	CATGTTGCGTTTTGTCCCTAAC
	RR491(R)	CCAAGTTGTTTTTAGAGAAAGTATCATATC
Cas9	RR656(F)	GGTGAGATCGTGTGGGATAAGG
	RR646(R)	GCACACAAGCTATTTATTTGACACACC

7.15 *Proteomics*

7.15.1 Protein extraction

Frozen plant material (~100mg) was ground using metal beads and a Tissue Lyser II (QIAGEN). The ground plant material was mixed in 1.5 volumes (150 µL) cold plant extraction buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% (v/v) glycerol, 0.1% (v/v) NP-40) with 1 mM PMSF and cOmplete™ Mini, EDTA-free Protease Inhibitor Cocktail (Roche; 1 tablet per 50 mL). The mixture was clarified by centrifuging at 14000rpm for 5 min at 4^{o}C. The supernatant was transferred to a clean 1.5 mL tube and the debris discarded. The extract was clarified again using the same conditions and the supernatant was transferred to a new tube and stored at -80^{o}C. The protein concentration was determined using the 2-D Quant Kit (GE) according to the manufacture's specifications.

7.15.2 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used for the size separation of proteins. 1.0 mm mini gels were made consisting of a 5% stacking phase over a 10% resolving phase.

5% Stacking SDS-gel – 5 mL

- MiliQ Water 3.5 mL
- 30% Acrylamide/Bis-acrylamide (37.5:1) 0.825 mL
- 1 M Tris HCl (pH 6.8) 0.625 mL
- 20% SDS 25 µL
- 10% Ammonium persulphate 50 µL
- N, N, N′, N′-tetramethylethylenediamine 5 µL

10% Resolving SDS-gel – 10 mL

- MiliQ Water 4.08 mL
- 30% Acrylamide/Bis-acrylamide (37.5:1) 3.3 mL
- 1 M Tris HCl (pH 6.8) 2.5 mL
- 20% SDS 50 µL
- 10% Ammonium persulphate 100 µL
- N, N, N′, N′-tetramethylethylenediamine 4 µL

Before loading 100 µg of the protein extracts were mixed with 4X Laemmli buffer (200 mM Tris-HCl pH6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 50% glycerol) to a 1X concentration and denatured by incubating at 95^{o}C for 5 min. PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) was used as the marker. SDS-PAGE was performed in running buffer (25 mM Tris-HCl pH8.3, 192 mM glycine, 3.5mM SDS). The gels were run at room temperature at 25 mA until samples had migrated through the stacking gel then increased to at 50 mA and run until the dye front had reached the lower rim of the gel.

7.15.3 Coomassie staining

SDS-gels were incubated in Coomassie staining solution (40% ethanol (v/v), 10% acetic acid (v/v), 0.1% (w/v) Coomassie Brillant Blue R250). The gels were then de-stained using destaining solution (20% methanol (v/v), 10% acetic acid (v/v)) for ~1 hour.

7.15.4 Protein detection

Slices of the gel corresponding to approximately 52 kDa (the molecular weight of UGT76E12 and UGT76E11) were cut and transferred to a 1.5 mL tube. Proteomic analysis was performed by Dr. Wolfgang Hoehenwarter using the following protocol (also provide by Dr. Hoehenwarter):

Proteins were in solution digested with trypsin and proteins were desalted as described by Majovsky and collaborators with some modifications (Majovsky et al., 2014). Dried peptides were dissolved in 5% acetonitrile, 0.1% trifluoric acid and 2 µg were injected into an EASY-nLC 1000 liquid chromatography system (Thermo Fisher Scientific). One to three micrograms of peptides were separated using liquid chromatography C18 reverse phase chemistry employing a 180 min gradient increasing from 5% to 40% acetonitrile in 0.1% FA, and a flow rate of 250 nL/min. Eluted peptides were electrosprayed on-line into a QExactive Plus mass spectrometer (Thermo Fisher Scientific). The spray voltage was 1.9 kV, the capillary temperature 275°C and the Z-Lens voltage 240 V. A full MS survey scan was carried out with chromatographic peak width set to 15 s, resolution 35,000, automatic

gain control (AGC) 1E+06 and a max injection time (IT) of 100 ms. MS/MS peptide sequencing was performed using a PRM scan strategy (without retention time scheduling) with HCD fragmentation containing 14 target peptide m/z on a list. MS scans with mass to charge ratios (m/z) between 200 and 1200 were acquired. MS/MS scans were acquired with resolution 17,500, AGC 2E+05, IT 100 ms, isolation width 1.6 m/z, normalized collision energy 27.

Peptides and proteins were identified using the Mascot software v2.5.0 (Matrix Science) linked to Proteome Discoverer v 2.1 (Thermo Fisher Scientific). A precursor ion mass error of 5 ppm and a fragment ion mass error of 0.02 Da were tolerated in searches of the TAIR10 database amended with common contaminants. Carbamidomethylation of cysteine (C) was set as fixed modification and oxidation of methionine (M) was tolerated as a variable modification. A spectrum (PSM), peptide and protein level false discovery rate (FDR) was calculated for all annotated PSMs, peptide groups and proteins based on the target-decoy database model and the percolator module. PSMs, peptide groups and proteins with q-values beneath the significance threshold of 0.01 for PSMs and peptides and 0.05 for proteins were considered identified.

7.16 *RNA stability assay*

RNA stability was determined as described previously by Fedak et al (*Fedak et al., 2016*). Briefly, *A. thaliana* Col-0 seedlings were grown on 1/2 MS (Murashige and Skoog Basal Medium) plates and after 11 d transferred to a 6-well-plate containing incubation buffer (1 mM PIPES pH 6.25, 1 mM trisodium citrate, 1 mM KCl, 15mM sucrose). After 30 min of incubation, 150 mg/l cordycepin (3'-deoxyadenosine) was added and vacuum-infiltrated (2x 5 min). Seedlings were collected at regular time-points every 40 min (0, 40, 80, 120 min) and frozen in liquid nitrogen. Total RNA was isolated, cDNA was synthesized and transcripts were analyzed by qRT-PCR. *EXPANSIN-LIKE1* (*Expansin L1*), a short-lived transcript, was used as a control. Ct values were normalized by the Ct value at time point 0 (Ct_0, formula 1) and plotted as degradation curves. Slope of this curve was used to calculate RNA half-lives ($t_{1/2,}$ formula 2).

$$Ct_n = (\ln(\frac{Ct}{Ct_0})) \times (-10) \qquad (1)$$

$$t_{1/2} = \frac{\ln(2)}{|slope|} \qquad (2)$$

Table 7.11 Primers for RNA stability assay

Gene	Primer	Sequence (5'→3')
EXPANSIN L1 (AT3G45970)	RR574(F)	CAAGTCGGTTCATCGCCAAATTGGG
	RR575(R)	GTATCCACCGGTTACTACGAACCTG
UGT76E12 (AT3G46660)	RR766(F)	GTCCAAGCTCCCTTGAAAGAAAC
	RR767(R)	CTAATGATGCAAACCGTGAAACTGG

7.17 *DNA methylation assay*

7.17.1 Growth conditions

Seedlings were grown on vertical plates (1.5% phytoagar) for 10 days after stratification under long-day conditions according to the general growth conditions explained in section **7.3**.

7.17.2 DNA methylation analysis

DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). The EpiMark Bisulfite Conversion Kit (New England Biolabs) was used to convert cytosine to uracil and prepare the sample for methylation analysis by PCR. The PCR was performed with EpiMark Hot Start *Taq* DNA Polymerase (New England Biolabs) and primers listed in **Table 7.12**. Two sets of primers were created, standard and converted. The converted primers share the same sequence as the standard, however, all cytosines were changed to thymine. The reactions were analyzed using agarose gel electrophoresis, products were extracted with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and cloning was performed with the pGEM-T Easy vector system (Promega). The plasmids were transformed into chemo competent TOP10 *E. coli* and grown on LB plates containing 0.5 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) and 80 µg/mL X-Gal. White colonies were selected and colony PCR using DreamTaq DNA polymerase (Thermo Fisher Scientific) was performed with M13 forward (-21) and reverse (-29) primers to confirm the presence of the inserts. Positive clones were grown in liquid media and plasmid preparations were quantified and sent to Eurofins Genomics for sequencing. Due to the high number of thymines and adenosines generated after bisulfite conversion, a "Power Read Upgrade" for bisulfite samples was required for sequencing. The sequences were analyzed using the webtool Quantification tool for Methylation Analysis (http://quma.cdb.riken.jp/) and the occurrence of cytosine methylation in each clone at each location was then graphed (Kumaki et al., 2008).

Table 7.12 Primers for DNA methylation assay

Gene	Primer	Type	Sequence (5'→3')
UGT76E12 (AT3G46660)	RR800(F)	Standard	F-GGAACATGCCAAACTTAGAGAGTAAAG
	RR801(R)		R-GCAGCTTCAGCAAAGTACATGAACTCAT
	RR802(F)	Converted (C→T)	F-GGAATATGTTAAATTTAGAGAGTAAAG
	RR803(R)		R-ACAACTTCAACAAAATACATAAACTCAT

7.18 *Chromatin Immunoprecipitation (ChIP)*

The ChIP assay was performed using the Plant ChIP kit (ab117137) according to manufacturer's specifications (abcam).

7.18.1 Growth conditions

Seedlings were grown for 10 days after stratification under long-day conditions according to the liquid culture conditions explained in section **7.4**. Prior to use, seedlings were washed 3 times in Milli-Q water and thoroughly dried to remove residual growth media.

7.18.2 Crosslinking

Cross-linking was performed as suggested by the Plant ChIP kit with approximately 1g plant material. To summarize, samples were submerged in 1% formaldehyde and vacuum infiltrated for 20 min. The cross-linking was quenched with 0.125 M Glycine, vacuum infiltrated again for 5 min and seedling were washed 3 times with ice-cold deionized water. The samples were thoroughly dried, frozen in liquid nitrogen, and stored at -80^{o}C.

7.18.3 Sonication

DNA was sheared by sonication using Sonopuls sonicator (Bandelin) and the TS 102 adaptor (Bandelin). The sonication was performed in 1.5mL micro tubes in an ice water bath. The intensity and cycling were optimized to obtain DNA fragments between 200-1000 bp. The sonicator was set at 60% power and set to cycle 5 times 15 sec ON and 45 sec OFF.

7.18.4 Immunoprecipitation

The IP reactions contained 100 µL of the sheared chromatin extract and 5 µL was used for the input DNA sample. Each reaction used 3 µg of the following antibodies: H3 (ab18521), H3K4me3 (ab8580), H3K36me3 (ab9050), H4K5ac (ab51997), H3K27ac (ab4729), and H3K27me3 (Millipore). All samples were eluted with 20 µL elution buffer.

7.18.5 Quantification and data analysis

The samples were quantified using qPCR with the Fast SYBR Green Mastermix (Thermo Fisher Scientific) and QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific). The reactions were set up and run as mentioned in section **7.10**. All the primers pairs used the standard annealing temperature of 60ºC excluding primers for location 5, which required a temperature of 64ºC (**Table 7.13**). The gene *ACT2* was used as a positive specificity control for H3K4me3 enrichment and *At4g03770* detected in samples immunoprecipitated with H3K9me2 was used as positive control for assessing kit functionality. The samples were diluted by 1/4 and 1 µL was used in the 10 µL qPCR reaction mix.

The "Ct method" was used for analysis. First, the dilution factor (DF) was calculated using formula (1). If the DF = 1/20 this translates to a 20X dilution.

$$Dilution\ factor = \frac{Volume_{Input}}{Volume_{IP}} \times \frac{Dilution_{Input}}{Dilution_{IP}} \quad (1)$$

The DF then needs to be converted into a "Ct equivalent" using formula (2) and solving for x. If dilution is 20X then x = log(20)/log(2).

$$2^x = Dilution \ \rightarrow\ x = \frac{\log\,(Dilution)}{\log\,(2)} \quad (2)$$

Next, %input can be calculated based on Ct values with formula (3)

$$\%Input = (Primer\ efficiency^{(Ct_{Input} - x - Ct_{IP})}) \times 100 \quad (3)$$

The average %Inputs were then graphed and compared between the genotypes.

Table 7.13 Primers used in ChIP analysis.

Gene	Location	Primer	Sequence (5'→3')
UGT76E12 (AT3G46660)	1	RR830(F)	GAAATTGTGACAAGATCCAACTCCATC
		RR831(R)	GGTTCTCACACTTGTGTACATTGATTC
	2	RR832(F)	CGAGTTTATCTCCTTTGGCACG
		RR833(R)	GAGGCTTCAAACAGAAAAAATTGTC
	3	RR834(F)	GACATGAAGCGAAATGCGGCAAG
		RR835(R)	GAATCATGTGGTGGATGGTGTTAC
	4	RR836(F)	GTTGTGACGAAGAGAGCCAAC
		RR837(R)	CCTGCTCAGCTTCGTGGTTC
	5	RR840(F)	GGGTGATCAAAAGGTGAACGC
		RR841(R)	GCTCTCTCGACCACTCCTC
	6	RR842(F)	GTGTCTACTAAAACACACAAAGCATC
		RR843(R)	CTGCAGGTGAGTTTTTAAGTGATGTTC
ACT2 (H3K4me3 enriched)	ACT2	RR849(F)	GGCGACTTGACAGAGAAGAAC
		RR849(R)	GAAAGAGCGGAAGAAGATGAGATTG
Transposon	At4g03770	RR855(F)	CATTTACGCGAATGTTTCAAGCAGC

(kit positive control)		RR856(R)	GATTTGCTTCGGATTTGCCTCG

7.19 *T-DNA insertion lines*

The T-DNA lines were identified from the SALK T-DNA Express database (http://signal.salk.edu/cgi-bin/tdnaexpress) and purchased from NASC. Lines *ugt76e11*^{D} (GK-783G02-025723) and *ugt76e12*^{D} (SALK_143394.54.75.X) were identified to have enhanced expression of *UGT76E11* and *UGT76E12*, respectively. Additionally, we observed nat-*ugt76e12*^{D} (SAILseq_444_E05.1) to have enhanced *NAT-UGT76E12* expression. Plant DNA was extracted using the extraction solution mention in section **7.14**. The genotyping PCR was performed using DreamTaq (Thermo Fisher Scientific) and the primers listed in **Table 7.14**. Homozygous lines containing the insertion were isolated and the exact location of the insert was confirmed by sequencing the PCR products at Eurofins Genomics.

Table 7.14 Primers for T-DNA genotyping

Line	Primer	Sequence (5'→3')
ugt76e11^{D}	RR441(F)	AGAGGAGCTCCAGGTGCTTAC
	RR442(R)	GTTGGCTCTCTTCGTCACAAC
	RR444(BP)	ATATTGACCATCATACTCATTGC
ugt76e12^{D}	RR439(F)	ATGGTGTTCGATGAAATGCTC
	RR440(R)	CTTACACTCTTTGGCTGCAGC
	RR443(BP)	ATTTTGCCGATTTCGGAAC
nat-ugt76e12^{D}	RR566(F)	GCCTCTGTTAAAAGTGGAGGC
	RR567(R)	GTATCGTTCATTTTCGCTTGC
	RR568(BP)	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC

7.20 *Time course experiments*

7.20.1 Growth conditions

Col-0 seedlings were grown on vertical plates (1.5% phytoagar) under long-day conditions according to the general growth conditions explained in section **7.3.** After 7 days, seedlings were transferred to 8 cm round pots with soil and grown for 45 days at 22^{o}C with 60% humidity in long-day conditions. Reporter gene lines were grown in similar conditions, however, seedlings were kept on plates until 25 days. For time points after 25 days, seedlings were transferred to soil at 7 days and grown until collection.

7.20.2 Collection and analysis

Samples for qRT-PCR analysis were collected at 14, 22, 25, 35, and 45 days after stratification. Pools of 12-15 whole rosettes for 14 day old plants and 3-4 whole rosettes for later time points were cut and frozen immediately in liquid nitrogen. Samples were homogenized and aliquots were taken for RNA extraction and cDNA synthesis. Transcripts were quantified using qRT-PCR as mentioned in section **7.10**.

Samples from the reporter gene lines were placed immediately in staining solution and incubated overnight at 37ºC. The protocol for GUS staining was followed as outlined in section **7.12**.

7.21 *Salt stress*

7.21.1 Growth Conditions

Samples meant for expression analysis were grown for 10 days after stratification under long-day conditions according to the liquid culture conditions explained in section **7.4**. Reporter gene lines were grown on vertical plates (1.5% phytoagar) for 10 days under long-day conditions according to the general growth conditions explained in section **7.3**.

7.21.2 Treatment

For samples grown in flasks, the experimental group was treated with 10 mL of media containing 1/2 MS, 1.5% (w/v) sucrose, and 2.75 M NaCl in order to reach final concentration of 250 mM. NaCl. The control group was treated with standard growth medium. The cultures were returned to the growth chamber and incubated for 6hrs. After, the seedlings were removed from the growth medium, thoroughly dried and quickly frozen in liquid nitrogen. Total RNA was extracted and used for cDNA synthesis. Expression levels of selected genes was analyzed by qRT-PCR using *UBC9* as a reference gene since *PP2AA3* was unstable in these experimental conditions.

The reporter gene lines were treated in a similar way as mentioned above, however, at 10 days seedlings were transferred to liquid medium containing 1/2 MS and 1.5% (w/v) sucrose at a pH of 5.6 with and without 250 mM NaCl. After treatment, seedlings were placed immediately in staining solution and incubated overnight at 37ºC. The protocol for GUS staining was followed as outlined in section **7.12**.

7.22 *B. cinerea infection assay*

7.22.1 A. thaliana growth

Seedlings were grown on round plates (0.8% phytoagar) for 10 days under 12hr light/dark conditions according to the general growth conditions explained in section **7.3**. After 10 days, seedlings were transferred to 8 cm round pots filled with soil in a tray and grown for 35 days at 22ºC with 60% humidity. For infection experiments performed on plates, seedlings were grown in similar conditions for 7 days in ½ MS sucrose free media.

7.22.2 B. cinerea growth

B. cinerea strain B05.10 was obtained from Dr. Lennart Eschen-Lippold and inoculated, 5 days before the planned infection time, onto a commercially purchased canned peach halve under sterile conditions. The inoculated peach was then incubated for 5 days under UV and white fluorescent light until spore development.

7.22.3 Spore preparation

A darker grey section of the infected peach was removed and placed into a 2 mL microtube containing 1 mL B5-Glc medium (3.16 g/L solid Gamborg B5 Medium with vitamins (Duchefa) and 2% glucose (w/v)). The spores were separated by rigorously vortexing and the suspension was filtered via nylon filter into a new 2 mL microtube. The spores were counted with a Fuchs-Rosenthal counting chamber (depth 0.200 mm, 0.0625 mm^2, Marienfeld) and the final required volume adjusted to $2x10^5$ spores/mL (5*spore number*1000*counting dilution = spores/mL). In order to initiate simultaneous spore germination, phosphate buffer at pH 6.4 to a final concentration of 10 mM was added and the suspension incubated at RT for 1 h.

7.22.4 Infection

35 day old plants were infected by inoculating 4 leaves per plant with 10µL of inoculum per leaf. Leaves at the same developmental stage were selected and the droplets were positioned toward the tip of the leaf avoiding the midrib. To avoid evaporation, the tray was then filled with a shallow amount of water and covered to maintain 100% humidity. The trays were carefully returned to the growth chamber and left for 48 hours. After 48 hpi, sets of 6 leaves of the same developmental stage were selected and photographed. A 3 mm cork-borer was used to make discs centered around the infected tissue. Discs were then transferred to a 2 mL microtube (each biological replicate contained 6 leaf discs) and immediately frozen

in liquid nitrogen. For infections performed on plates, 10µL of inoculum were directly pipetted on individual seedlings than were grown for an additional 24 hours. Control samples were treated with a spore-free inoculum. The on-plate inoculated seedlings were collected for RNA extraction and expression levels quantified using qRT-PCR

7.22.5 Lesion area and DNA quantification

Lesion area was measured using ImageJ with an in-image scale as reference. The regions considered to be part of the lesion were areas with obvious symptoms of necrosis. Symptoms include leaf discoloration (darker color) and missing tissue (usually cause by handling).

Lesion discs were kept frozen and pulverized to a fine powder with a 4mm metal bead in a TissueLyser II (QIAGEN) (x2 at 25Hz for 1 min). Extraction of DNA was performed with DNeasy Plant Mini Kit (QIAGEN) using a QIAcube (QIAGEN) machine according to manufacturer's instructions. A plasmid containing a gene from *Solanum tuberosum* (*StNOX*) not found in *A. thaliana* was added during the lysis (step 2) and used as an external control. The DNA was eluted once with 50 µL of the provided elution buffer and stored at –80^{o}C. The concentrations of DNA extracts were measured using TECAN Infinite® 200 spectrophotometer and the NanoQuant Plate™. Absorbance measurements were made at 260 nm and 280 nm to determine DNA concentration and purity. Extractions with a 260/280 ratio of approximately 1.8 were considered accurate and contaminant-free. The average concentration of two technical replicates was used to ensure correct values. Before the qPCR, DNA samples were diluted to a concentration of 1ng/µL in a 96 well plate and the volume of DNA required was noted.

B.cinerea DNA was quantified via qPCR using the standard curve method with 2X EvaGreen(ROX) (Bio&SELL) mastermix according to the manufacturer's instructions. The QuantStudio^{TM} 5 Real-Time PCR System (Thermo Fisher Scientific) was set to custom cycle outlined in **Table 7.15**. The reactions were carried out in 10 µL using 1.5 µL of DNA [1 ng/µL] as template and pipetted by a QIAgility (Qiagen). The *B. cinerea CUTINASE A* (*BcCUTA*) gene and the *StNOX* gene from the supplemented plasmid were used as targets (**Table 7.16**). A standard curve was set up with a starting concentration of 10 ng/µL *B.cinerea* DNA and 1ng/µL control plasmid. The dilutions were made with a serial factor of 1:10 to a total of 5 points.

In order to calculate the quantity of *B. cinerea* DNA normalized by *StNOX* the following equations were used: First, we calculated the theoretical quantity of *StNOX*,

assuming a zero-loss during extraction (if 20 ng plasmid was added and the sample was eluted in 50 µL the zero loss concentration would be 0.4 ng/µL), present in the 1ng/µL diluted sample.

$$StNOX \text{ in sample (ng)} = \text{DNA for dilution (µL)} \times \text{zero loss } (\frac{\text{ng}}{\text{µL}})$$

Second, we calculated the theoretical amount of *StNOX* in the PCR reaction (in our case we used 1.5 µL).

$$StNOX \text{ in PCR (ng)} = StNOX \text{ in sample(ng)} \times \text{PCR template volume(µL)}$$

Third, we determined the expected Ct_{StNOX} value based on the calculated amount in the PCR and the slope plus the Y-intercept from the standard curve.

$$\text{Expected } Ct_{StNOX} = (\text{slope}_{StNOX} \times \log(StNOX \text{ in PCR})) + \text{Y intercept}_{StNOX}$$

Fourth, the Ct_{BcCUTA} value was corrected based on the expected and measured Ct_{StNOX}. This step is required to correct for variations in efficiencies of the extractions.

$$\text{Corrected } Ct_{BcCUTA} = \left(\frac{\text{Expected } Ct_{StNOX}}{\text{Measured } Ct_{StNOX}}\right) \times \text{Measured } Ct_{BcCUTA}$$

Finally, the quantity of *B.cinerea* DNA was calculated based on the standard curve.

$$\text{Quantity } B.cinerea\ DNA = 10^{\frac{\text{Corrected } Ct_{BcCUTA} - \text{Y intercept}_{BcCUTA}}{\text{slope}_{BcCUTA}}}$$

Table 7.15 EvaGreen PCR protocol.

Cycle step	Temperature	Time	Cycle number
Initial incubation	95ºC	15min	1x
Denaturing	95ºC	15sec	40
Annealing	64ºC	40sec	

Table 7.16 Primers used to quantify *B. cinerea* DNA.

Gene	Primer	Sequence (5'→3')
BcCUTA	RR692(F)	GAAGAGAAATGGAAAATGGTGAG
BcCUTA	RR693(R)	AGCCTTATGTCCCTTCCCTTG
StNOX	RR694(F)	AATTCGCGTAGCTCATCTTTAGTCA
StNOX	RR695(R)	GAGTTTGCTGAGGAATTATTTGATGCT

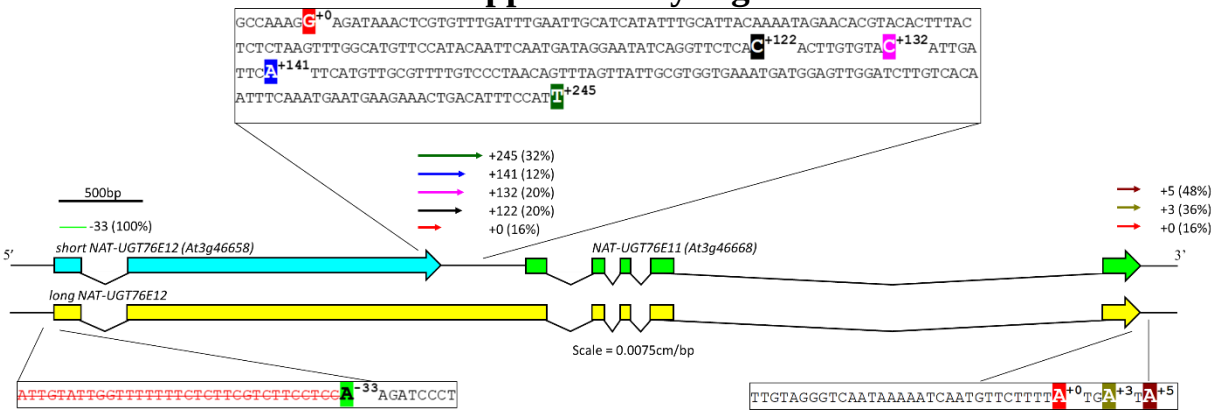
7.22.6 Graphing and statistics

The data was presented in fold change in comparison to *WT*. Regarding DNA quantity, each sample was divided by the mean of all *WT* samples per experiment. The normalized values were used for statistical analysis and graphed. Lesion size was handled in a similar fashion, but individual lesion areas were divided by the means of the *WT* samples of each biological replicate.

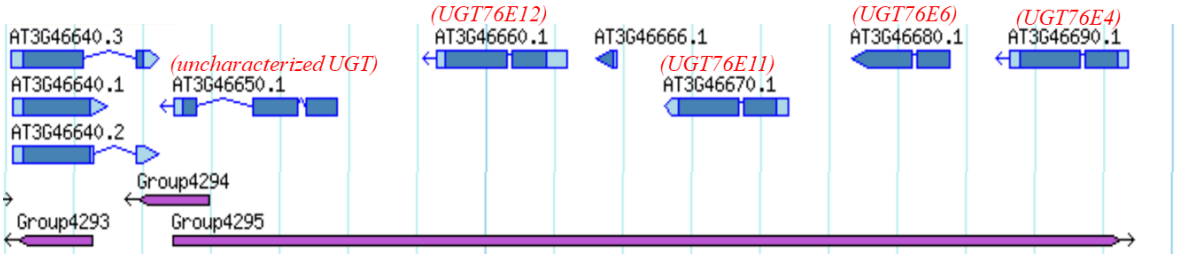
7.23 *Statistics and graphs*

All statistics and graphs were done using Microsoft Excel (Microsoft) and Prism8 (Graphpad Software).

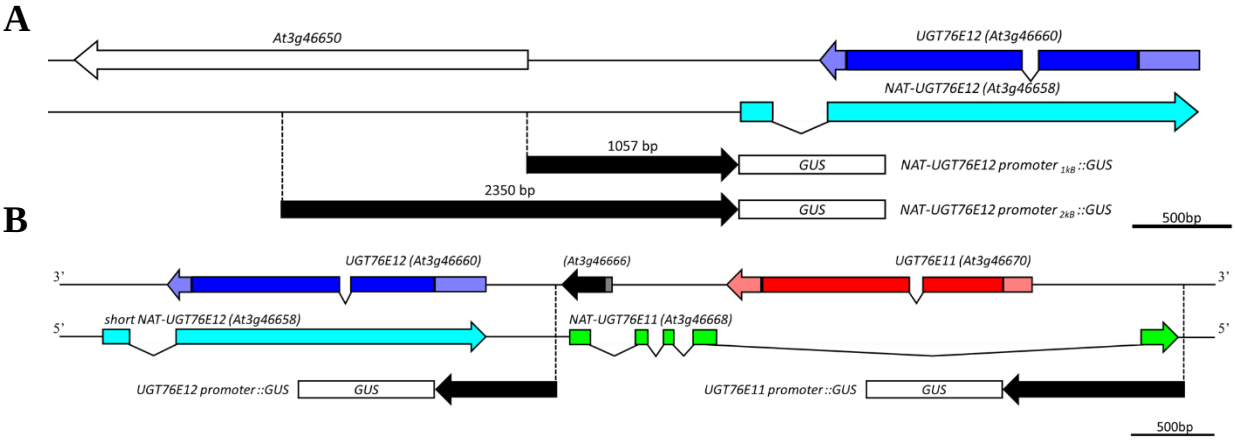
8 Supplementary Figures



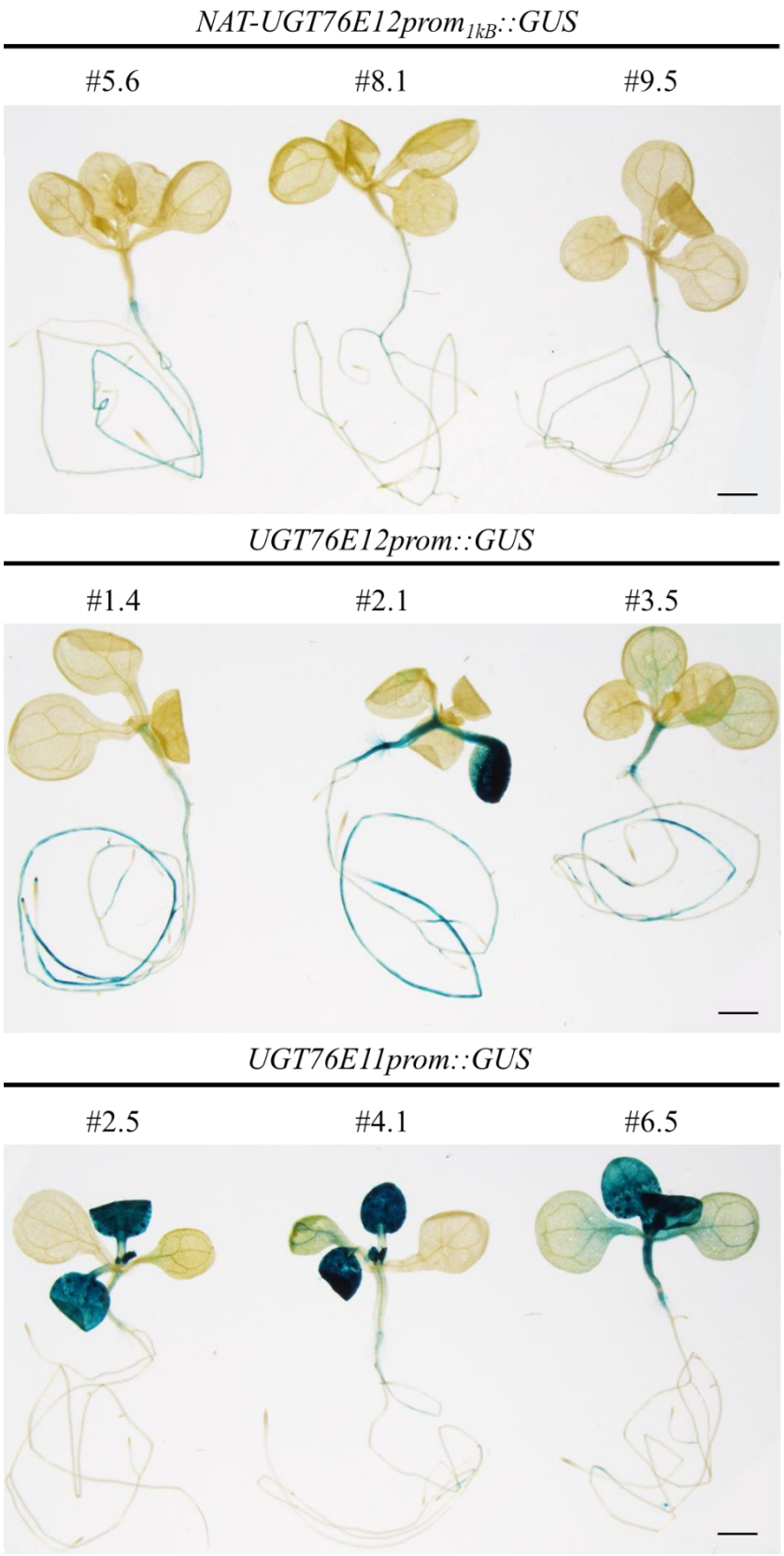
Supplementary Figure 8.1 *NAT-UGT76E12* 5' and 3' ends determined by RACE. The line represents 5' RACE and arrows represent 3' RACE results showing length and frequency of the different observed ends of *NAT-UGT76E12 in vivo*. The insets show the varying transcriptional initiation and termination sites. Bases in red with a strikethrough indicate the sequence of the end annotated in TAIR10 but not observed in the RACE experiments. Number of clones analyzed n=22 – 25 for each RACE experiment.



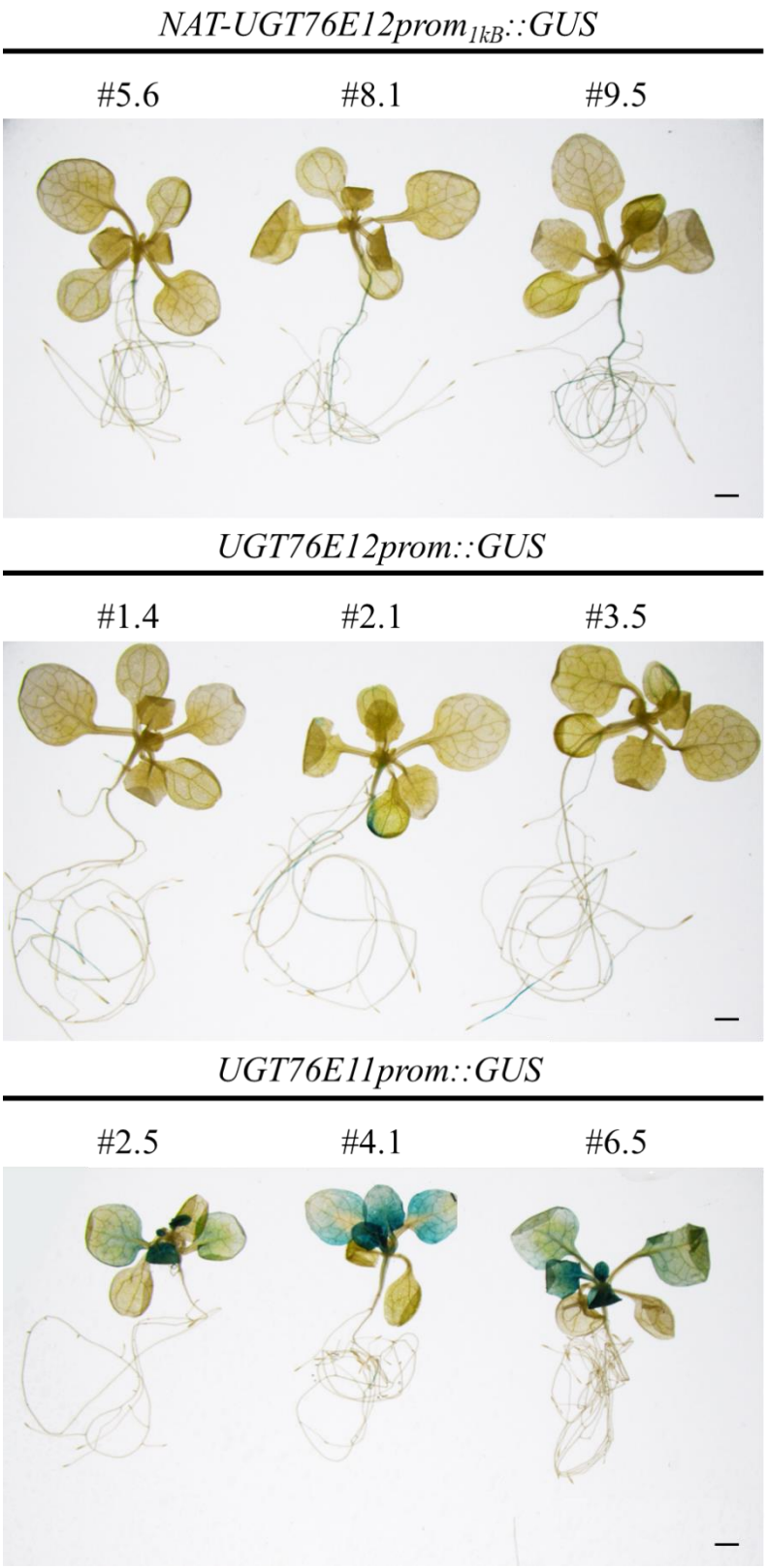
Supplementary Figure 8.2 Scheme summarizing tiling array data from the Plant long noncoding RNA Database (PLncDB). Locus of *UGT76E11* and *UGT76E12* (Matsui et al., 2010). Group4295 is a proposed NAT spanning 13.8kB overlapping *UGT76E11*, *UGT76E12*, *UGT76E4*, *UGT76E6*, and an uncharacterized UGT.



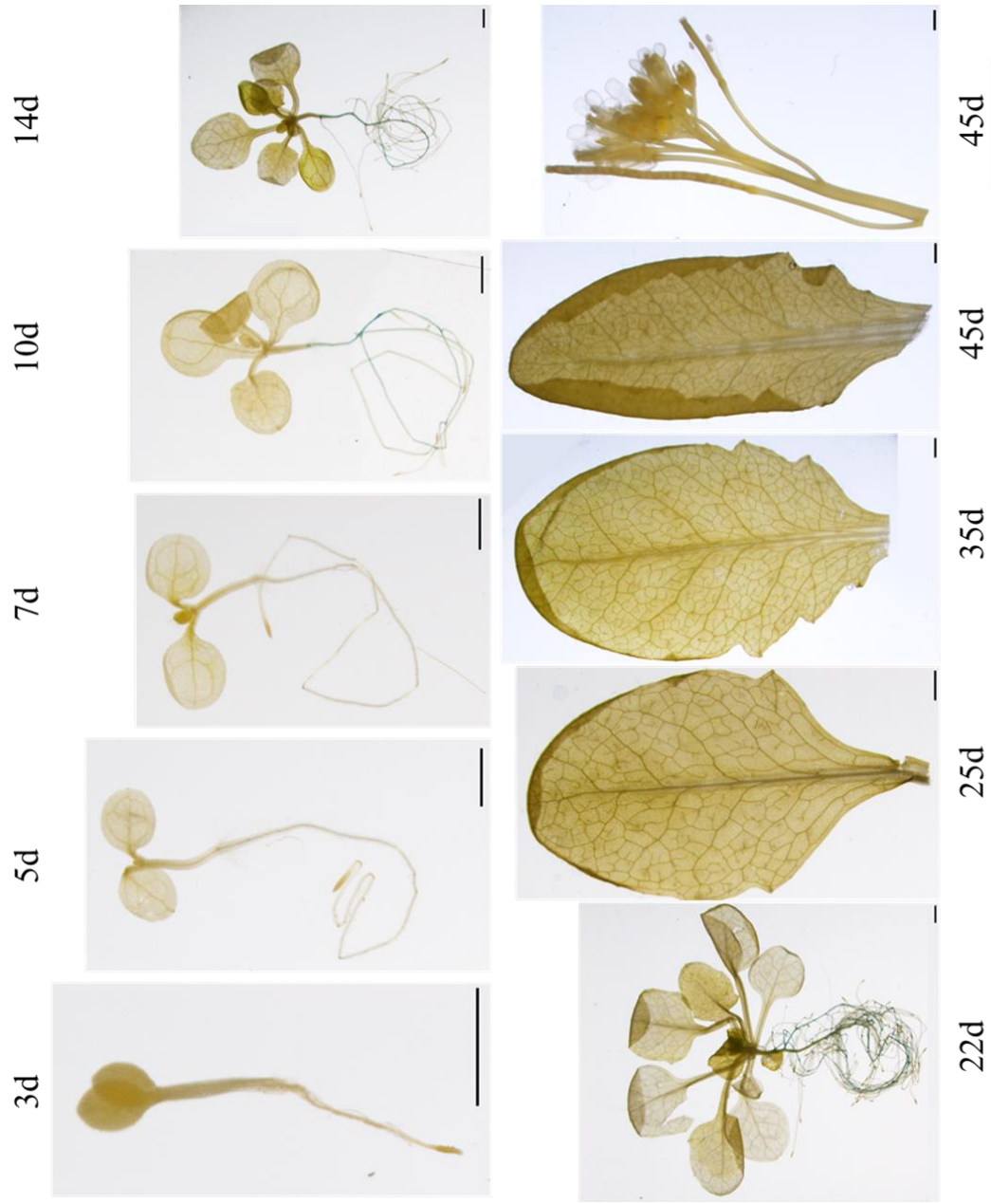
Supplementary Figure 8.3 *Promoter::GUS* constructs. ***A***, Schematic outlining the promoter region of *NAT-UGT76E12* used in the promoter reporter lines. Two independent constructs were generated including 1057 and 2350 bp upstream of transcription initiation site (simplified 1kB and 2kB, respectively). ***B***, Reporter lines for *UGT76E11* and *UGT76E12* were generated using 1090 bp and 731 bp upstream of the start codon, respectively.



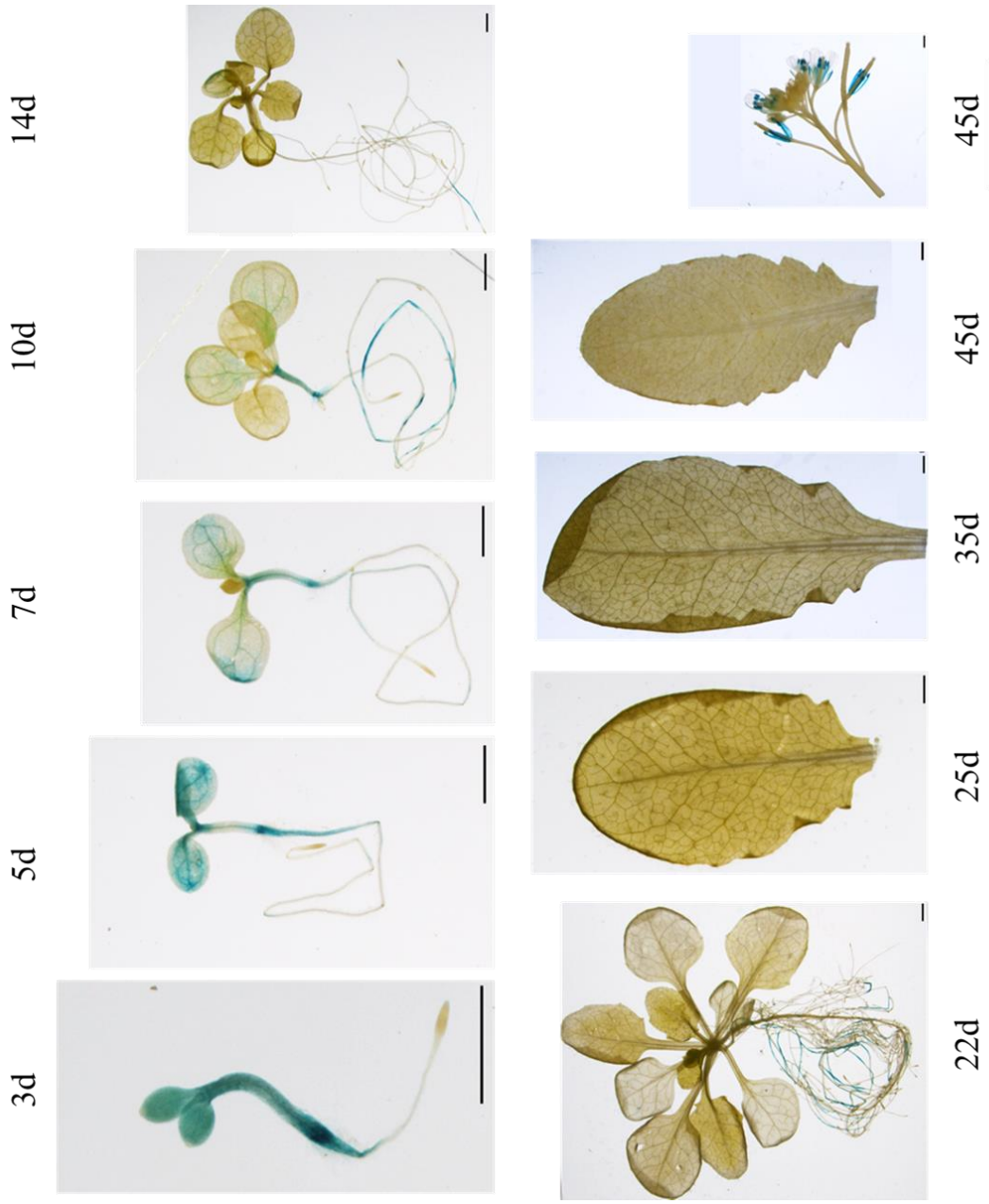
Supplementary Figure 8.4 Histochemical staining of three independent *promoter::GUS* reporter gene lines showing promoter activity in 10 d old seedlings. Seedlings were grown on vertical plates in long-day conditions. Staining was performed overnight at 37ºC. Scale 1.0 cm.



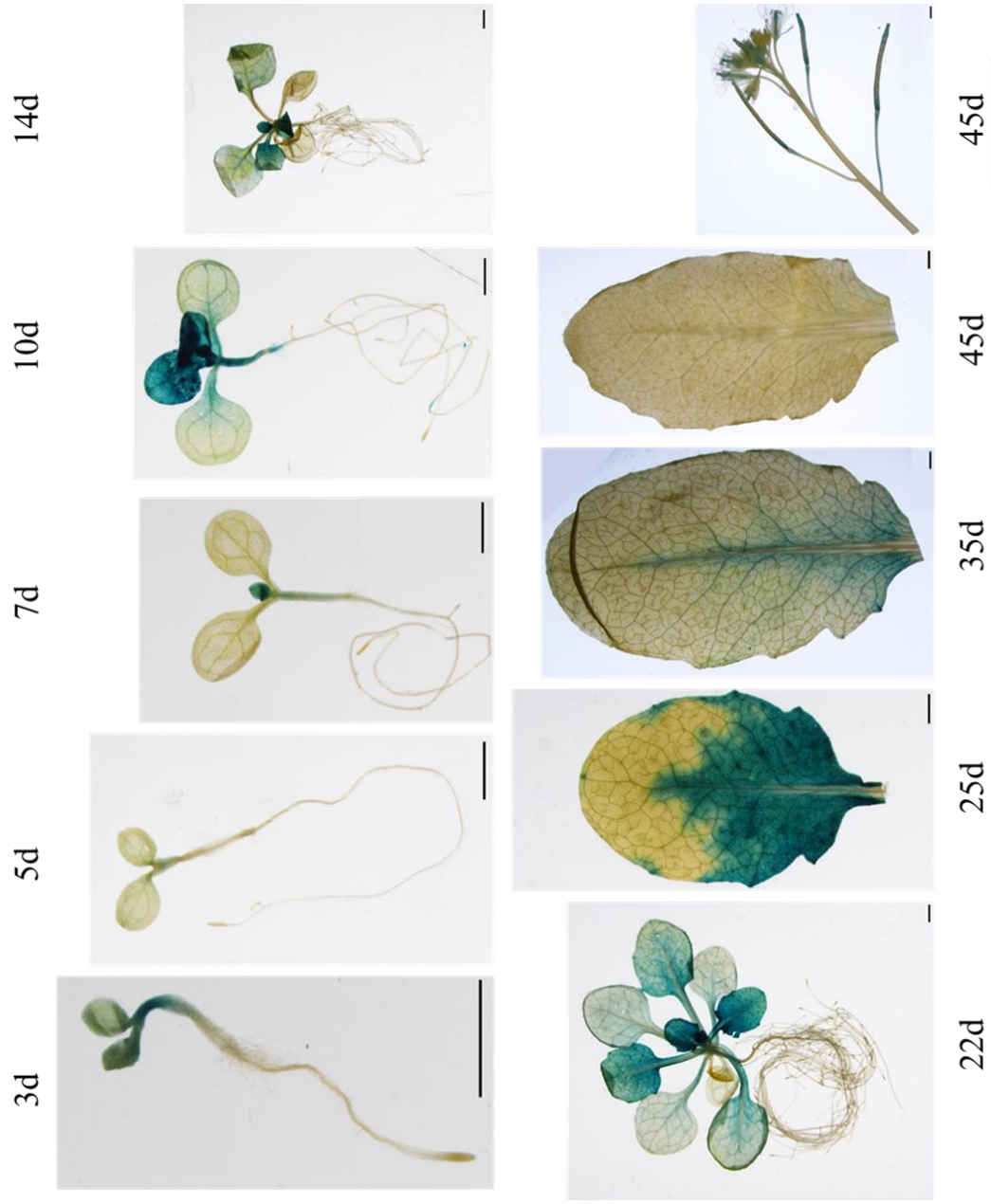
Supplementary Figure 8.5 Histochemical staining of three independent *promoter::GUS* reporter gene lines showing promoter activity in 14 day old seedlings. Seedlings were grown on vertical plates in long-day conditions. Staining was performed overnight at 37ºC. Scale 1.0 cm



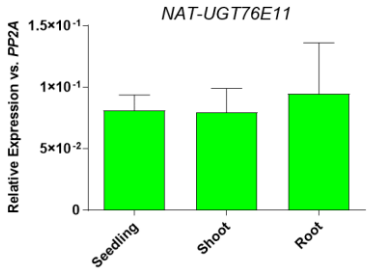
Supplementary Figure 8.6 Histochemical staining of *NAT-UGT76E12prom$_{1kB}$::GUS* at various stages of development. Histochemical staining at 3, 5, 7, 10, 14 22, 25, 35, 45 days after stratification (das). Seedlings collected at 3 to 22 das were grown on vertical plates. Later time points were collected from plants grown on soil. Line #9.5 was used as a representative from three independent lines. Staining was performed overnight at 37ºC. Scale 1.0 cm.



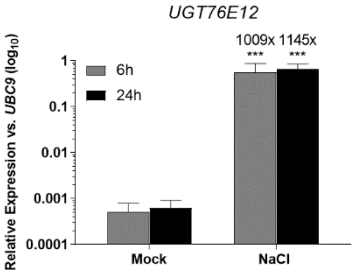
Supplementary Figure 8.7 Histochemical staining of *UGT76E12prom::GUS* at various stages of development. Histochemical staining at 3, 5, 7, 10, 14 22, 25, 35, 45 das. Seedlings collected at 3 to 22 das were grown on vertical plates. Later time points were collected from plants grown on soil. Line #3.5 was used as a representative from three independent lines. Staining was performed overnight at 37ºC. Scale 1.0 cm



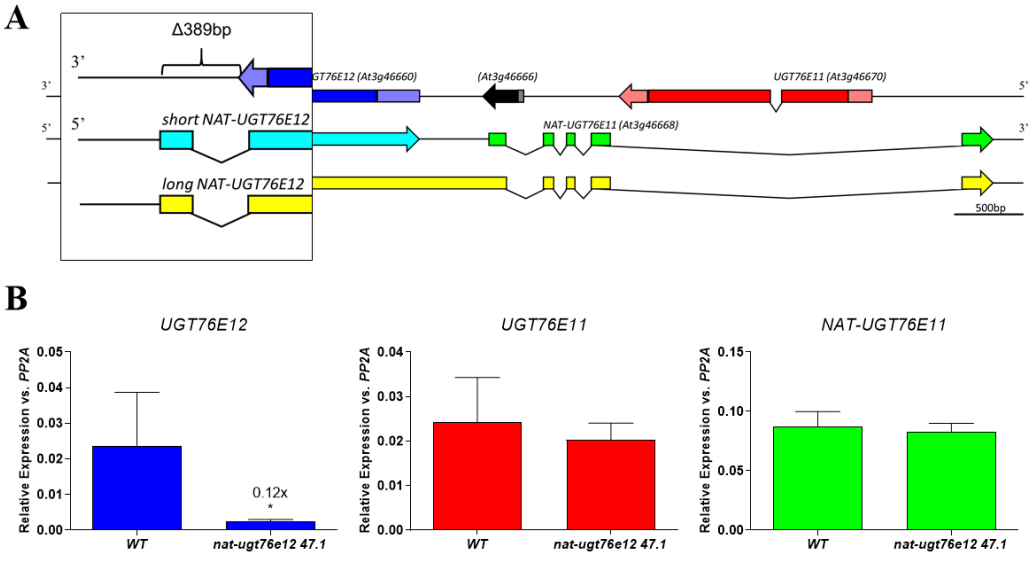
Supplementary Figure 8.8 Histochemical staining of *UGT76E11prom::GUS* at various stages of development. Histochemical staining at 3, 5, 7, 10, 14 22, 25, 35, 45 das. Seedlings collected at 3 to 22 das were grown on vertical plates. Later time points were collected from plants grown on soil. Line #6.5 was used as a representative from three independent lines. Staining was performed overnight at 37ºC. Scale 1.0 cm



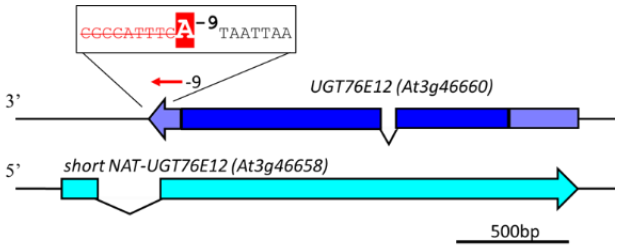
Supplementary Figure 8.9 Expression levels of *NAT-UGT76E12* in seedlings. qRT-PCR of 10 day old seedlings comparing expression in shoots, roots, and complete seedlings. Data are mean ± SD from 3 biological replicates (n=3) from 1 of 3 independent experiments with similar results.



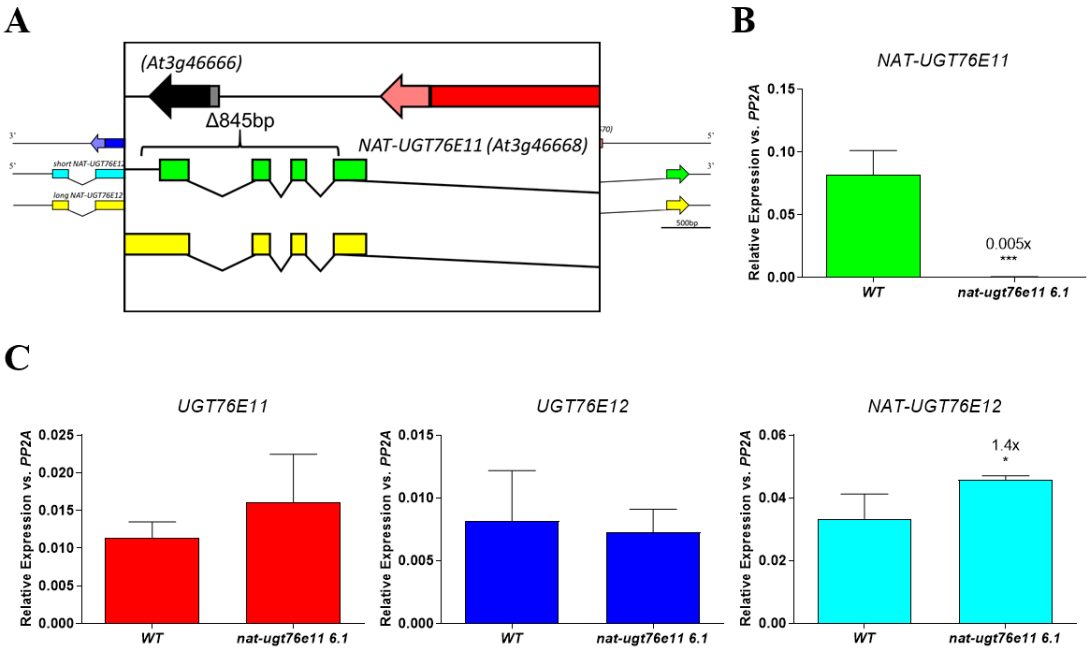
Supplementary Figure 8.10 Effect of 6 and 24 h of salt stress on *UGT76E12* expression levels. qRT-PCR of 10 day old seedlings grown in liquid media after 6 and 24 hours of 250mM NaCl treatment. Mock samples were supplemented with media without salt. Data are mean ± SD from 3 biological replicates (n=3) from 1 of 2 independent experiments with similar results. Values above bars indicates fold change in comparison to *WT*. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001.



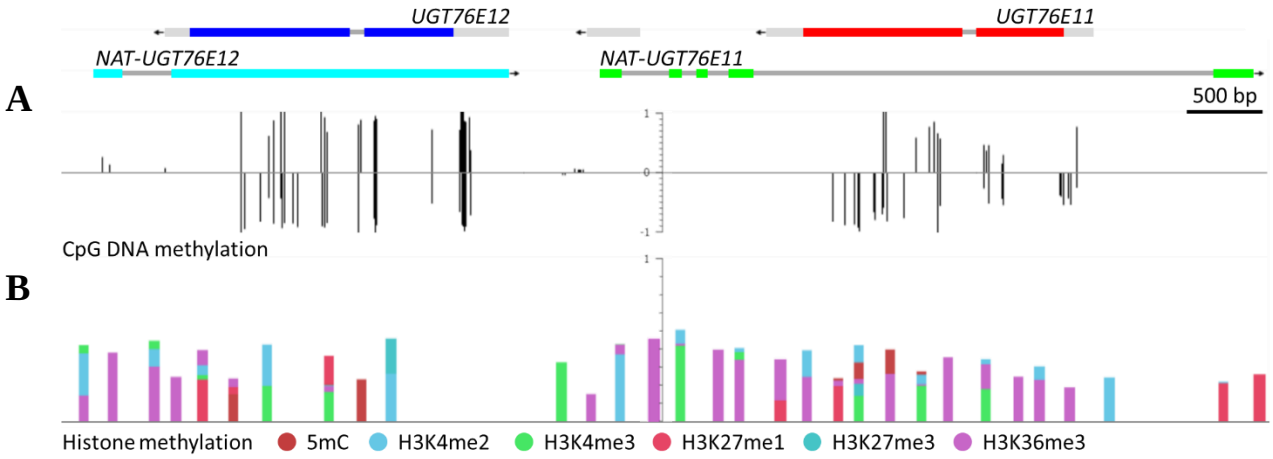
Supplementary Figure 8.11 Additional line showing the effect of *nat-ugt76e12* knockout on gene expression. ***A***, Schematic of the *UGT76E12* genomic region showing the 389 bp CRISPR/Cas9 mediated deletion of the first exon of *NAT-UGT76E12*. ***B,*** qRT-PCR showing transcript levels in *nat-ugt76e12 47.1* vs. *WT*. Total RNA extracted from 10 day old seedlings and cDNA synthesized using gene specific primers. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. One-tailed t-test; *p<0.05.



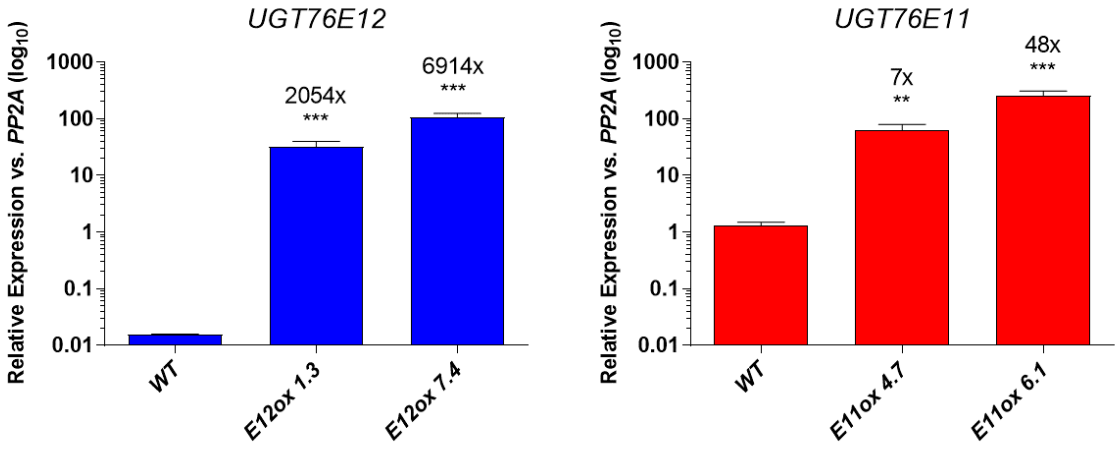
Supplementary Figure 8.12 Schematic representation of the *UGT76E12* 3' end based on RACE. The arrow represents 3' RACE results showing the observed end of *UGT76E12 in vivo*. Results were obtained from direct sequencing of PCR product. The inset shows the transcription termination site which is located 9 nt upstream as reported in TAIR10. Bases in red with a strikethrough indicate the sequence of the end annotated in TAIR10 but not observed in the RACE.



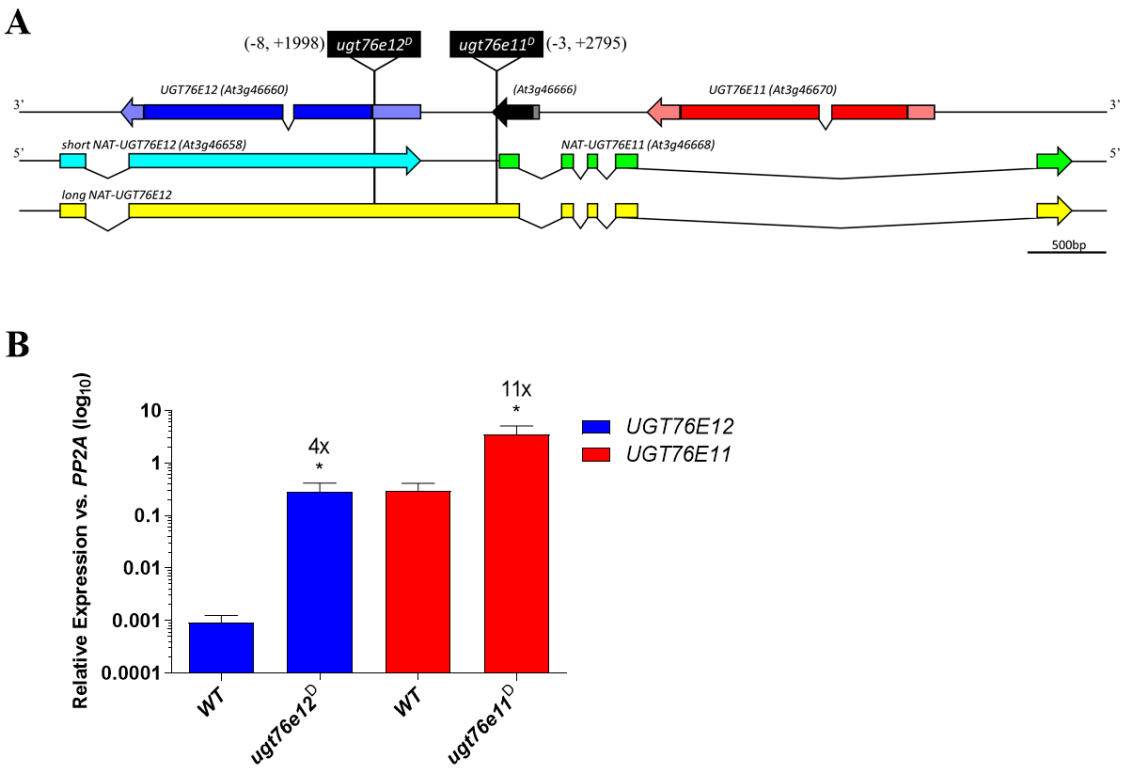
Supplementary Figure 8.13 Effect of *nat-ugt76e11* knockout on gene expression levels. ***A,*** Schematic of an 845 bp CRISPR/Cas9 mediated deletion of the first three exons of *NAT-UGT76E11*. ***B,*** qRT-PCR showing the level of *NAT-UGT76E11* expression. ***C,*** qRT-PCR comparing transcript levels in *nat-ugt76e11 6.1* vs. *WT*. Total RNA extracted from 10 day old seedlings. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001.



Supplementary Figure 8.14 Epigenomic mapping of the chromatin state at the *UGT76E11* and *UGT76E12* locus. ***A,*** Bisulfite-Seq of CpG DNA methylation (Stroud, Greenberg, Feng, Bernatavichute, & Jacobsen, 2013) and ***B,*** ChIP-on-Chip assay (below) (Roudier et al., 2011). ChIP-on-ChIP used antibodies against 5-methylcytosine (5mC), H3K4me2, H3K4me3, H3K27me1, H3K27me3, and H3K36me3. Figure modified from CoGe database (Lyons & Freeling, 2008). UTRs (light grey) and introns (dark gray).



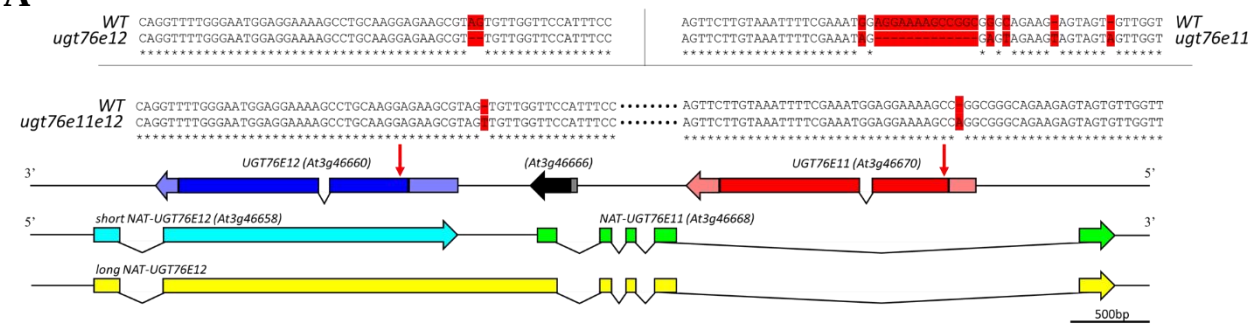
Supplementary Figure 8.15 Levels of *UGT76E11* and *UGT76E12* in the overexpression lines. Level of overexpression in comparison to *WT* in 2 independent overexpression lines for each gene. Data are mean ± SD from one experiment with 3 biological replicates (n=3). Values above bars indicates fold change in comparison to *WT*. One-tailed T-test *p<0.05, **p<0.01, ***p<0.001.



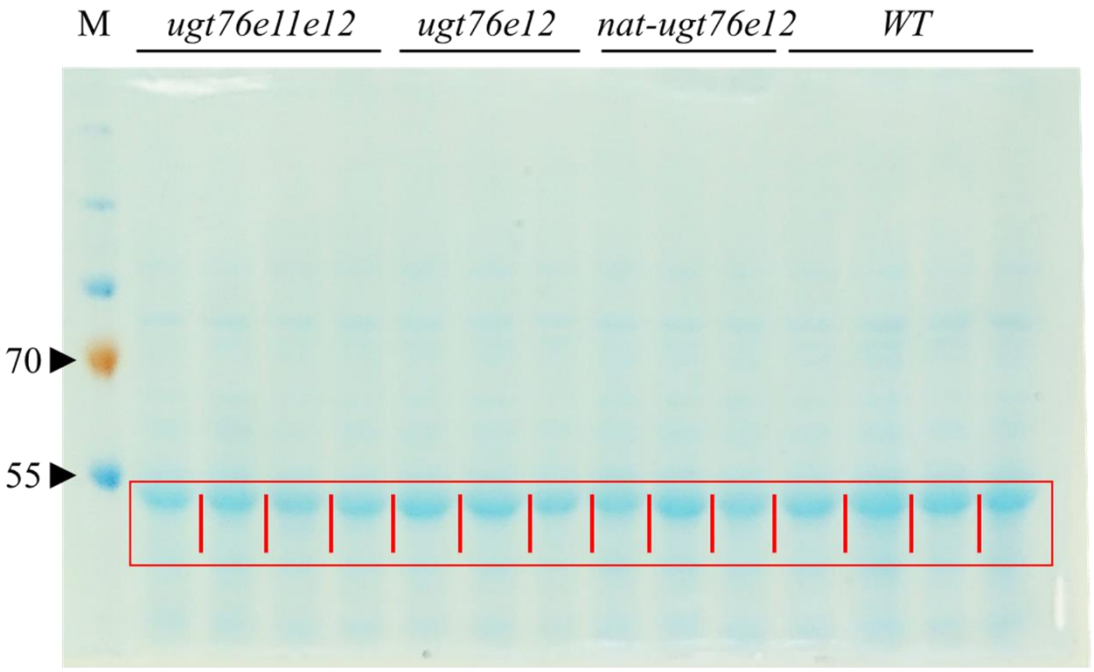
Supplementary Figure 8.16 Location of T-DNA insertion mutants. ***A,*** Two independent homozygous insertional mutant lines *ugt76e11*^{D} and *ugt76e12*^{D} result in an upregulation of *UGT76E11* and *UGT76E12*, respectively. The position of the insertion is noted in parenthesis. Positive and negative values indicate upstream and downstream of the start codon or to the TIS for protein-coding genes and long non-coding RNA genes, respectively (+1 is the first nucleotide). The insertion in *ugt76e11*^{D} is -3 bp upstream of *NAT-UGT76E11* and + 2795bp downstream of *NAT-UGT76E12*. The insertion in *ugt76e12*^{D} is -8 bp upstream of *UG76E12* and +1998bp downstream of *NAT-UGT76E12*. ***B***, Expression levels in the T-DNA lines showing increased basal levels of *UGT76E11* and *UGT76E12*. Data are mean ± SD from 3 biological replicates (n=3) from 1 of 2 independent experiments with similar results. Values above bars indicates fold change in comparison to *WT*. One-tailed T-test *p<0.05, **p<0.01, ***p<0.001.

Supplementary Figures

A



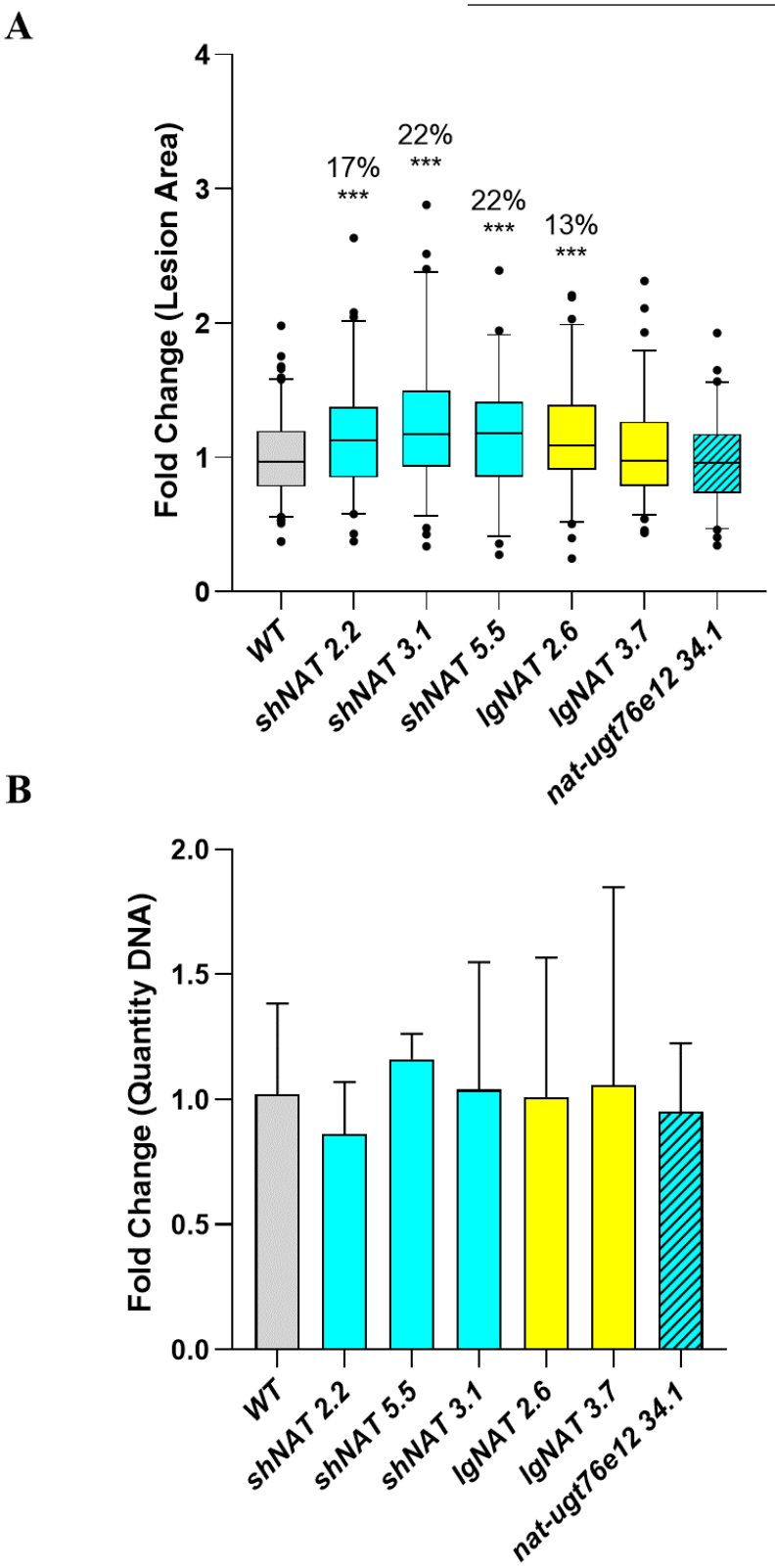
B



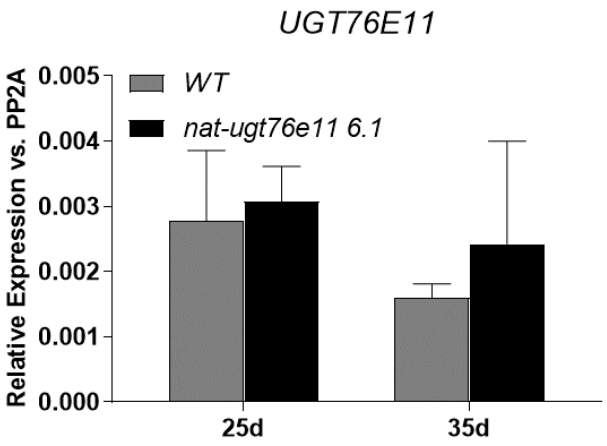
C

Sample (left to right)	UGT76E11	UGT76E12	UGT78D1	UGT71D1
ugt76e11e12 6.2	-	-	+	+
ugt76e11e12 6.2	-	-	+	+
ugt76e11e12 6.2	-	-	+	+
ugt76e11e12 6.2	-	-	+	+
ugt76e12 23.6	+	-	+	+
ugt76e12 23.6	+	-	+	+
ugt76e12 23.6	+	-	+	+
nat-ugt76e12 34.1	+	-	+	+
nat-ugt76e12 34.1	+	+	+	+
nat-ugt76e12 34.1	+	+	+	+
WT	+	+	+	+
WT	+	+	+	+
WT	+	+	+	+
WT	+	+	+	+

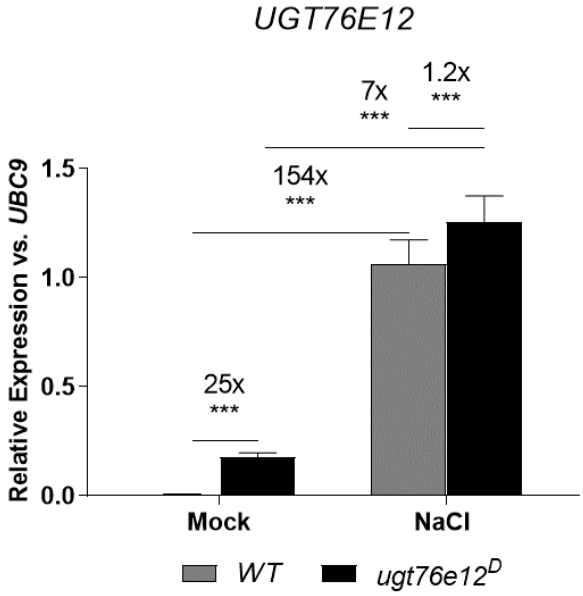
Supplementary Figure 8.17 CRISPR/Cas knockout of protein coding genes. ***A***, Schematic representation of target locus for single and double mutants. Red arrows indicate cutting site. Sequence alignments comparing *WT* to mutant lines. Indels are highlighted in red for each line all resulting in translational knockouts. ***B***, Preparative SDS-PAGE of protein extracted from10 day old seedlings grown in liquid media after 6 hrs of treatment with 250mM NaCl. Red boxes highlights regions excised from gel that were used for proteomic analysis. M, Protein molecular weight marker (kDa) Arrowheads on the left indicate the proteins of 55 and 70 kDa ***C***, Results from proteomic analysis using LC/MS/MS. (+) and (-) indicate the presence or absence of the protein, respectively. UGT78D1 and UGT76D1 were used as positive controls of protein detection.



Supplementary Figure 8.18 Effects of altered *NAT-UGT76E12* expression on susceptibility to *B. cinerea* infection. *A,* Fold change of lesion area 48hpi expressed as a percentage compared with the *WT* control. Boxplot indicates median and outliers are 0-5% and 95-100% from compilation of 3 independent experiments (n=48-144). ***B,*** Fold change of quantity of *B. cinerea* DNA measured using qPCR expressed as a percentage compared with the *WT* control. Data are mean ± SD from 3 independent experiments (n=8-24). Values above bars indicates fold change in comparison to *WT*. One-tailed t-test *p<0.05, **p<0.01, ***p<0.001.



Supplementary Figure 8.19 *UGT76E11* expression levels in *WT* and *nat-ugt76e11 6.1*. qRT-PCR of 25 and 35 day old plants from *WT* and *nat-ugt76e11 6.1* lines. Data are mean ± SD from one experiment with 4 biological replicates (n=4). Each sample contains pools of 4-12 rosettes with removed stems.



Supplementary Figure 8.20 Effect of salt stress on *UGT76E12* expression levels in the *ugt76e12^D* line. qRT-PCR of *WT* and *ugt76e12^D* 10 day old seedlings grown in liquid media after 6 hours of 250mM NaCl treatment. Seedlings supplemented with media without salt were used as controls (mock). Data are mean ± SD from 4 biological replicates (n=4) from 1 of 2 independent experiments with similar results. One-tailed t-test; *p<0.05, **p<0.01, ***p<0.001.

9 References

AbuQamar, S., Moustafa, K., & Tran, L. S. (2017). Mechanisms and strategies of plant defense against Botrytis cinerea. *Crit Rev Biotechnol, 37*(2), 262-274. doi:10.1080/07388551.2016.1271767

Agrios, G. (2005). Plant Pathology, ISBN: 9780120445653. 2005. In: Academic Press.-Madigan MT, Martinko JM, Stahl DA, Clark DP. Brock Biology ….

Ariel, F., Romero-Barrios, N., Jegu, T., Benhamed, M., & Crespi, M. (2015). Battles and hijacks: noncoding transcription in plants. *Trends Plant Sci, 20*(6), 362-371. doi:10.1016/j.tplants.2015.03.003

Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, M., & Matzke, A. J. (2002). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J, 21*(24), 6832-6841.

Bach, D. H., & Lee, S. K. (2018). Long noncoding RNAs in cancer cells. *Cancer Lett, 419*, 152-166. doi:10.1016/j.canlet.2018.01.053

Bari, R., & Jones, J. D. (2009). Role of plant hormones in plant defence responses. *Plant Mol Biol, 69*(4), 473-488. doi:10.1007/s11103-008-9435-0

Baurle, I., & Dean, C. (2006). The timing of developmental transitions in plants. *Cell, 125*(4), 655-664. doi:10.1016/j.cell.2006.05.005

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., & Woodcock, C. L. (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*(24), 14173-14178. doi:10.1073/pnas.95.24.14173

Brazier-Hicks, M., Gershater, M., Dixon, D., & Edwards, R. (2018). Substrate specificity and safener inducibility of the plant UDP-glucose-dependent family 1 glycosyltransferase super-family. *Plant Biotechnol J, 16*(1), 337-348. doi:10.1111/pbi.12775

Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci U S A, 99 Suppl 4*, 16491-16498. doi:10.1073/pnas.162371599

Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet, 10*(5), 295-304. doi:10.1038/nrg2540

Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol, 27*, 207-216. doi:10.1016/j.pbi.2015.08.003

Chen, M., & Penfield, S. (2018). Feedback regulation of COOLAIR expression controls seed dormancy and flowering time. *Science, 360*(6392), 1014-1017. doi:10.1126/science.aar7361

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J, 89*(4), 789-804. doi:10.1111/tpj.13415

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., . . . Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature, 452*(7184), 215-219. doi:10.1038/nature06745

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science, 322*(5909), 1845-1848. doi:10.1126/science.1162228

Crisp, P. A., Smith, A. B., Ganguly, D. R., Murray, K. D., Eichten, S. R., Millar, A. A., & Pogson, B. J. (2018). RNA Polymerase II Read-Through Promotes Expression of Neighboring Genes in SAL1-PAP-XRN Retrograde Signaling. *Plant Physiol, 178*(4), 1614-1630. doi:10.1104/pp.18.00758

Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci U S A, 111*(45), 16160-16165. doi:10.1073/pnas.1419030111

Daviere, J. M., & Achard, P. (2013). Gibberellin signaling in plants. *Development, 140*(6), 1147-1151. doi:10.1242/dev.087650

De Miccolis Angelini, R. M., Rotolo, C., Masiello, M., Gerin, D., Pollastro, S., & Faretra, F. (2014). Occurrence of fungicide resistance in populations of Botryotinia fuckeliana (Botrytis cinerea) on table grape and strawberry in southern Italy. *Pest Manag Sci, 70*(12), 1785-1796. doi:10.1002/ps.3711

Dean, J. V., & Delaney, S. P. (2008). Metabolism of salicylic acid in wild-type, ugt74f1 and ugt74f2 glucosyltransferase mutants of Arabidopsis thaliana. *Physiol Plant, 132*(4), 417-425. doi:10.1111/j.1399-3054.2007.01041.x

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., . . . Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol, 13*(4), 414-430. doi:10.1111/j.1364-3703.2011.00783.x

Doolittle, W. F. (2013). Is junk DNA bunk? A critique of ENCODE. *Proc Natl Acad Sci U S A, 110*(14), 5294-5300. doi:10.1073/pnas.1221376110

El Oirdi, M., El Rahman, T. A., Rigano, L., El Hadrami, A., Rodriguez, M. C., Daayf, F., . . . Bouarab, K. (2011). Botrytis cinerea manipulates the antagonistic effects between immune pathways to promote disease development in tomato. *Plant Cell, 23*(6), 2405-2421. doi:10.1105/tpc.111.083394

Elad, Y., Pertot, I., Prado, A. M. C., & Stewart, A. (2016). Plant hosts of Botrytis spp. In *Botrytis–the fungus, the pathogen and its management in agricultural systems* (pp. 413-486): Springer.

Ellis, C., Karafyllidis, I., Wasternack, C., & Turner, J. G. (2002). The Arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses. *Plant Cell, 14*(7), 1557-1566. doi:10.1105/tpc.002022

Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., . . . Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med, 14*(7), 723-730. doi:10.1038/nm1784

Fedak, H., Palusinska, M., Krzyczmonik, K., Brzezniak, L., Yatusevich, R., Pietras, Z., . . . Swiezewski, S. (2016). Control of seed dormancy in Arabidopsis by a cis-acting noncoding antisense transcript. *Proc Natl Acad Sci U S A, 113*(48), E7846-E7855. doi:10.1073/pnas.1608827113

Finnegan, E. J., & Dennis, E. S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. *Nucleic Acids Res, 21*(10), 2383-2388. doi:10.1093/nar/21.10.2383

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev, 67*(1), 16-37, table of contents. doi:10.1128/mmbr.67.1.16-37.2003

George Thompson, A. M., Iancu, C. V., Neet, K. E., Dean, J. V., & Choe, J. Y. (2017). Differences in salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from Arabidopsis thaliana. *Sci Rep, 7*, 46629. doi:10.1038/srep46629

Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature, 470*(7333), 284-288. doi:10.1038/nature09701

Greenberg, M. V., Ausin, I., Chan, S. W., Cokus, S. J., Cuperus, J. T., Feng, S., . . . Jacobsen, S. E. (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics, 6*(3), 344-354.

Haroth, S., Feussner, K., Kelly, A. A., Zienkiewicz, K., Shaikhqasem, A., Herrfurth, C., & Feussner, I. (2019). The glycosyltransferase UGT76E1 significantly contributes to 12-O-glucopyranosyl-jasmonic acid formation in wounded Arabidopsis thaliana leaves. *J Biol Chem, 294*(25), 9858-9872. doi:10.1074/jbc.RA119.007600

Hawkes, E. J., Hennelly, S. P., Novikova, I. V., Irwin, J. A., Dean, C., & Sanbonmatsu, K. Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. *Cell Rep, 16*(12), 3087-3096. doi:10.1016/j.celrep.2016.08.045

He, X. J., Chen, T., & Zhu, J. K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res, 21*(3), 442-465. doi:10.1038/cr.2011.23

Henriques, R., Wang, H., Liu, J., Boix, M., Huang, L. F., & Chua, N. H. (2017). The antiphasic regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytol, 216*(3), 854-867. doi:10.1111/nph.14703

Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol, 19*(5), 327-341. doi:10.1038/nrm.2017.130

Heo, J. B., Lee, Y. S., & Sung, S. (2013). Epigenetic regulation by long noncoding RNAs in plants. *Chromosome Res, 21*(6-7), 685-693. doi:10.1007/s10577-013-9392-6

Heo, J. B., & Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science, 331*(6013), 76-79. doi:10.1126/science.1197349

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science, 308*(5718), 118-120. doi:10.1126/science.1106910

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P., & Bradbury, E. M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem, 268*(1), 305-314.

Hou, X., Lee, L. Y., Xia, K., Yan, Y., & Yu, H. (2010). DELLAs modulate jasmonate signaling via competitive binding to JAZs. *Dev Cell, 19*(6), 884-894. doi:10.1016/j.devcel.2010.10.024

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., . . . Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics, 2008*, 420747. doi:10.1155/2008/420747

Huang, S. C., & Ecker, J. R. (2018). Piecing together cis-regulatory networks: insights from epigenomics studies in plants. *Wiley Interdiscip Rev Syst Biol Med, 10*(3), e1411. doi:10.1002/wsbm.1411

Huang, Y., & Carmichael, G. G. (1996). Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol, 16*(4), 1534-1542.

Hughes, J., & Hughes, M. A. (1994). Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. *DNA Seq, 5*(1), 41-49.

Husar, S., Berthiller, F., Fujioka, S., Rozhon, W., Khan, M., Kalaivanan, F., . . . Poppenberger, B. (2011). Overexpression of the UGT73C6 alters brassinosteroid

glucoside formation in Arabidopsis thaliana. *BMC Plant Biol, 11*, 51. doi:10.1186/1471-2229-11-51

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., & Poirier, Y. (2013). A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell, 25*(10), 4166-4182. doi:10.1105/tpc.113.116251

Jackson, J. P., Lindroth, A. M., Cao, X., & Jacobsen, S. E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature, 416*(6880), 556-560. doi:10.1038/nature731

Jackson, R. G., Kowalczyk, M., Li, Y., Higgins, G., Ross, J., Sandberg, G., & Bowles, D. J. (2002). Over-expression of an Arabidopsis gene encoding a glucosyltransferase of indole-3-acetic acid: phenotypic characterisation of transgenic lines. *Plant J, 32*(4), 573-583.

Jiang, D., Kong, N. C., Gu, X., Li, Z., & He, Y. (2011). Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet, 7*(3), e1001330. doi:10.1371/journal.pgen.1001330

Joh, R. I., Palmieri, C. M., Hill, I. T., & Motamedi, M. (2014). Regulation of histone methylation by noncoding RNAs. *Biochim Biophys Acta, 1839*(12), 1385-1394. doi:10.1016/j.bbagrm.2014.06.006

Johnson, L. M., Du, J., Hale, C. J., Bischof, S., Feng, S., Chodavarapu, R. K., . . . Jacobsen, S. E. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature, 507*(7490), 124-128. doi:10.1038/nature12931

Kapitonov, D., & Yu, R. K. (1999). Conserved domains of glycosyltransferases. *Glycobiology, 9*(10), 961-978. doi:10.1093/glycob/9.10.961

Karimi, M., De Meyer, B., & Hilson, P. (2005). Modular cloning in plant cells. *Trends Plant Sci, 10*(3), 103-105. doi:10.1016/j.tplants.2005.01.008

Karimi, M., Inze, D., & Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci, 7*(5), 193-195.

Kim, D. H., & Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Dev Cell, 40*(3), 302-312 e304. doi:10.1016/j.devcel.2016.12.021

Kim, D. H., Xi, Y., & Sung, S. (2017). Modular function of long noncoding RNA, COLDAIR, in the vernalization response. *PLoS Genet, 13*(7), e1006939. doi:10.1371/journal.pgen.1006939

Kumaki, Y., Oda, M., & Okano, M. (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Res, 36*(Web Server issue), W170-175. doi:10.1093/nar/gkn294

Langlois-Meurinne, M., Gachon, C. M., & Saindrenan, P. (2005). Pathogen-responsive expression of glycosyltransferase genes UGT73B3 and UGT73B5 is necessary for resistance to Pseudomonas syringae pv tomato in Arabidopsis. *Plant Physiol, 139*(4), 1890-1901. doi:10.1104/pp.105.067223

Law, J. A., Du, J., Hale, C. J., Feng, S., Krajewski, K., Palanca, A. M., . . . Jacobsen, S. E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature, 498*(7454), 385-389. doi:10.1038/nature12178

Lee, D. Y., Hayes, J. J., Pruss, D., & Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell, 72*(1), 73-84.

Leon-Reyes, A., Du, Y., Koornneef, A., Proietti, S., Korbes, A. P., Memelink, J., . . . Ritsema, T. (2010). Ethylene signaling renders the jasmonate response of

Arabidopsis insensitive to future suppression by salicylic Acid. *Mol Plant Microbe Interact, 23*(2), 187-197. doi:10.1094/MPMI-23-2-0187

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell, 128*(4), 707-719. doi:10.1016/j.cell.2007.01.015

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*(16), 1501-1507. doi:10.1016/j.cub.2005.07.029

Li, Q., Yu, H. M., Meng, X. F., Lin, J. S., Li, Y. J., & Hou, B. K. (2018). Ectopic expression of glycosyltransferase UGT76E11 increases flavonoid accumulation and enhances abiotic stress tolerance in Arabidopsis. *Plant Biol (Stuttg), 20*(1), 10-19. doi:10.1111/plb.12627

Lim, E. K., Ashford, D. A., Hou, B., Jackson, R. G., & Bowles, D. J. (2004). Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol Bioeng, 87*(5), 623-631. doi:10.1002/bit.20154

Lim, E. K., Doucet, C. J., Li, Y., Elias, L., Worrall, D., Spencer, S. P., . . . Bowles, D. J. (2002). The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J Biol Chem, 277*(1), 586-592. doi:10.1074/jbc.M109287200

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., . . . Jacobsen, S. E. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J, 23*(21), 4286-4296. doi:10.1038/sj.emboj.7600430

Liu, C., Lu, F., Cui, X., & Cao, X. (2010). Histone methylation in higher plants. *Annu Rev Plant Biol, 61*, 395-420. doi:10.1146/annurev.arplant.043008.091939

Liu, J., Bai, G., Zhang, C., Chen, W., Zhou, J., Zhang, S., . . . Zhu, J. K. (2011). An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. *Cell Res, 21*(12), 1691-1700. doi:10.1038/cr.2011.173

Liu, X., Luo, M., & Wu, K. (2012). Epigenetic interplay of histone modifications and DNA methylation mediated by HDA6. *Plant Signal Behav, 7*(6), 633-635. doi:10.4161/psb.19994

Logemann, E., Birkenbihl, R. P., Ulker, B., & Somssich, I. E. (2006). An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. *Plant Methods, 2*, 16. doi:10.1186/1746-4811-2-16

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature, 389*(6648), 251-260. doi:10.1038/38444

Lyons, E., & Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J, 53*(4), 661-673. doi:10.1111/j.1365-313X.2007.03326.x

Majovsky, P., Naumann, C., Lee, C. W., Lassowskat, I., Trujillo, M., Dissmeyer, N., & Hoehenwarter, W. (2014). Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. *J Proteome Res, 13*(10), 4246-4258. doi:10.1021/pr500164j

Matsui, A., Ishida, J., Morosawa, T., Okamoto, M., Kim, J. M., Kurihara, Y., . . . Seki, M. (2010). Arabidopsis tiling array analysis to identify the stress-responsive genes. *Methods Mol Biol, 639*, 141-155. doi:10.1007/978-1-60761-702-0_8

Mattick, J. S., Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mehler, M. F. (2009). RNA regulation of epigenetic processes. *Bioessays, 31*(1), 51-59. doi:10.1002/bies.080099

Mattick, J. S., & Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol, 22*(1), 5-7. doi:10.1038/nsmb.2942

Matzke, M. A., Kanno, T., & Matzke, A. J. (2015). RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol, 66*, 243-267. doi:10.1146/annurev-arplant-043014-114633

Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet, 15*(6), 394-408. doi:10.1038/nrg3683

Meech, R., Hu, D. G., McKinnon, R. A., Mubarokah, S. N., Haines, A. Z., Nair, P. C., . . . Mackenzie, P. I. (2019). The UDP-Glycosyltransferase (UGT) Superfamily: New Members, New Functions, and Novel Paradigms. *Physiol Rev, 99*(2), 1153-1222. doi:10.1152/physrev.00058.2017

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet, 10*(3), 155-159. doi:10.1038/nrg2521

Monfort, A., & Wutz, A. (2017). Progress in understanding the molecular mechanism of Xist RNA function through genetics. *Philos Trans R Soc Lond B Biol Sci, 372*(1733). doi:10.1098/rstb.2016.0368

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology, 38*(1), 23-38. doi:10.1038/npp.2012.112

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., & Fukuhara, T. (2014). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. *Nucleic Acids Res, 42*(3), 1845-1856. doi:10.1093/nar/gkt1077

Noguchi, T., Fujioka, S., Choe, S., Takatsuto, S., Yoshida, S., Yuan, H., . . . Tax, F. E. (1999). Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. *Plant Physiol, 121*(3), 743-752. doi:10.1104/pp.121.3.743

Onodera, Y., Haag, J. R., Ream, T., Costa Nunes, P., Pontes, O., & Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell, 120*(5), 613-622. doi:10.1016/j.cell.2005.02.007

Osmani, S. A., Bak, S., & Moller, B. L. (2009). Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry, 70*(3), 325-347. doi:10.1016/j.phytochem.2008.12.009

Paquette, S., Moller, B. L., & Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry, 62*(3), 399-413.

Pikaard, C. S., & Mittelsten Scheid, O. (2014). Epigenetic regulation in plants. *Cold Spring Harb Perspect Biol, 6*(12), a019315. doi:10.1101/cshperspect.a019315

Piskurewicz, U., Jikumaru, Y., Kinoshita, N., Nambara, E., Kamiya, Y., & Lopez-Molina, L. (2008). The gibberellic acid signaling repressor RGL2 inhibits Arabidopsis seed germination by stimulating abscisic acid synthesis and ABI5 activity. *Plant Cell, 20*(10), 2729-2745. doi:10.1105/tpc.108.061515

Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., Schuhmacher, R., Krska, R., . . . Adam, G. (2003). Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from Arabidopsis thaliana. *J Biol Chem, 278*(48), 47905-47914. doi:10.1074/jbc.M307552200

Poppenberger, B., Fujioka, S., Soeno, K., George, G. L., Vaistij, F. E., Hiranuma, S., . . . Bowles, D. (2005). The UGT73C5 of Arabidopsis thaliana glucosylates

brassinosteroids. *Proc Natl Acad Sci U S A, 102*(42), 15253-15258. doi:10.1073/pnas.0504279102

Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol, 16*(3), 190-202. doi:10.1038/nrm3943

Qian, S., Lv, X., Scheid, R. N., Lu, L., Yang, Z., Chen, W., . . . Du, J. (2018). Dual recognition of H3K4me3 and H3K27me3 by a plant histone reader SHL. *Nat Commun, 9*(1), 2425. doi:10.1038/s41467-018-04836-y

Rademacher, T. W., Parekh, R. B., & Dwek, R. A. (1988). Glycobiology. *Annu Rev Biochem, 57*, 785-838. doi:10.1146/annurev.bi.57.070188.004033

Rosa, S., Duncan, S., & Dean, C. (2016). Mutually exclusive sense-antisense transcription at FLC facilitates environmentally induced gene repression. *Nat Commun, 7*, 13031. doi:10.1038/ncomms13031

Ross, J., Li, Y., Lim, E., & Bowles, D. J. (2001). Higher plant glycosyltransferases. *Genome Biol, 2*(2), REVIEWS3004.

Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., . . . Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J, 30*(10), 1928-1938. doi:10.1038/emboj.2011.103

Sotelo-Silveira, M., Chavez Montes, R. A., Sotelo-Silveira, J. R., Marsch-Martinez, N., & de Folter, S. (2018). Entering the Next Dimension: Plant Genomes in 3D. *Trends Plant Sci, 23*(7), 598-612. doi:10.1016/j.tplants.2018.03.014

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature, 403*(6765), 41-45. doi:10.1038/47412

Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V., & Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell, 152*(1-2), 352-364. doi:10.1016/j.cell.2012.10.054

Swiezewski, S., Liu, F., Magusin, A., & Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature, 462*(7274), 799-802. doi:10.1038/nature08618

Tanny, J. C. (2014). Chromatin modification by the RNA Polymerase II elongation complex. *Transcription, 5*(5), e988093. doi:10.4161/21541264.2014.988093

Tariq, M., & Paszkowski, J. (2004). DNA and histone methylation in plants. *Trends Genet, 20*(6), 244-251. doi:10.1016/j.tig.2004.04.005

Tian, Y., Zheng, H., Zhang, F., Wang, S., Ji, X., Xu, C., . . . Ding, Y. (2019). PRC2 recruitment and H3K27me3 deposition at FLC require FCA binding of COOLAIR. *Sci Adv, 5*(4), eaau7246. doi:10.1126/sciadv.aau7246

To, T. K., Kim, J. M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., . . . Seki, M. (2011). Arabidopsis HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLoS Genet, 7*(4), e1002055. doi:10.1371/journal.pgen.1002055

Van der Does, D., Leon-Reyes, A., Koornneef, A., Van Verk, M. C., Rodenburg, N., Pauwels, L., . . . Pieterse, C. M. (2013). Salicylic acid suppresses jasmonic acid signaling downstream of SCFCOI1-JAZ by targeting GCC promoter motifs via transcription factor ORA59. *Plant Cell, 25*(2), 744-761. doi:10.1105/tpc.112.108548

van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., . . . Amon, A. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell, 150*(6), 1170-1181. doi:10.1016/j.cell.2012.06.049

Verma, V., Ravindran, P., & Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol, 16*, 86. doi:10.1186/s12870-016-0771-y

Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane- Robinson, C., Allis, C. D., & Workman, J. L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J, 15*(10), 2508-2518.

Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T., & Steitz, J. A. (2015). Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell, 59*(3), 449-461. doi:10.1016/j.molcel.2015.06.016

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T. A., . . . Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc Natl Acad Sci U S A, 114*(40), E8362-E8371. doi:10.1073/pnas.1711120114

Vilborg, A., & Steitz, J. A. (2017). Readthrough transcription: How are DoGs made and what do they do? *RNA Biol, 14*(5), 632-636. doi:10.1080/15476286.2016.1149680

Voet, D., Voet, J. G., & Pratt, C. W. (2016). *Fundamentals of biochemistry : life at the molecular level* (5th edition. ed.). Hoboken, NJ: John Wiley & Sons.

von Saint Paul, V., Zhang, W., Kanawati, B., Geist, B., Faus-Kessler, T., Schmitt-Kopplin, P., & Schaffner, A. R. (2011). The Arabidopsis glucosyltransferase UGT76B1 conjugates isoleucic acid and modulates plant defense and senescence. *Plant Cell, 23*(11), 4124-4145. doi:10.1105/tpc.111.088443

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., & Chua, N. H. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res, 24*(3), 444-453. doi:10.1101/gr.165555.113

Wang, H. V., & Chekanova, J. A. (2017). Long Noncoding RNAs in Plants. *Adv Exp Med Biol, 1008*, 133-154. doi:10.1007/978-981-10-5203-3_5

Wang, J., Meng, X., Dobrovolskaya, O. B., Orlov, Y. L., & Chen, M. (2017). Non-coding RNAs and Their Roles in Stress Response in Plants. *Genomics Proteomics Bioinformatics, 15*(5), 301-312. doi:10.1016/j.gpb.2017.01.007

Wang, Y., Luo, X., Sun, F., Hu, J., Zha, X., Su, W., & Yang, J. (2018). Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat Commun, 9*(1), 3516. doi:10.1038/s41467-018-05829-7

Weiberg, A., Wang, M., Lin, F. M., Zhao, H., Zhang, Z., Kaloshian, I., . . . Jin, H. (2013). Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science, 342*(6154), 118-123. doi:10.1126/science.1239705

Wierzbicki, A. T., Haag, J. R., & Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell, 135*(4), 635-648. doi:10.1016/j.cell.2008.09.035

Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet, 41*(5), 630-634. doi:10.1038/ng.365

Williamson, B., Tudzynski, B., Tudzynski, P., & van Kan, J. A. (2007). Botrytis cinerea: the cause of grey mould disease. *Mol Plant Pathol, 8*(5), 561-580. doi:10.1111/j.1364-3703.2007.00417.x

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev, 23*(13), 1494-1504. doi:10.1101/gad.1800909

Wunderlich, M., Gross-Hardt, R., & Schoffl, F. (2014). Heat shock factor HSFB2a involved in gametophyte development of Arabidopsis thaliana and its expression is

controlled by a heat-inducible long non-coding antisense RNA. *Plant Mol Biol, 85*(6), 541-550. doi:10.1007/s11103-014-0202-0

Xing, D., & Li, Q. Q. (2011). Alternative polyadenylation and gene expression regulation in plants. *Wiley Interdiscip Rev RNA, 2*(3), 445-458. doi:10.1002/wrna.59

Xu, C., Tian, J., & Mo, B. (2013). siRNA-mediated DNA methylation and H3K9 dimethylation in plants. *Protein Cell, 4*(9), 656-663. doi:10.1007/s13238-013-3052-7

Yang, M., Fehl, C., Lees, K. V., Lim, E. K., Offen, W. A., Davies, G. J., . . . Davis, B. G. (2018). Functional and informatics analysis enables glycosyltransferase activity prediction. *Nat Chem Biol, 14*(12), 1109-1117. doi:10.1038/s41589-018-0154-9

Yin, Y., Zhao, Y., Wang, J., Liu, C., Chen, S., Chen, R., & Zhao, H. (2007). antiCODE: a natural sense-antisense transcripts database. *BMC Bioinformatics, 8*, 319. doi:10.1186/1471-2105-8-319

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., . . . Ecker, J. R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell, 126*(6), 1189-1201. doi:10.1016/j.cell.2006.08.003

Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., & Qi, Y. (2018). Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat Commun, 9*(1), 5056. doi:10.1038/s41467-018-07500-7

Zhao, Z., Wu, X., Ji, G., Liang, C., & Li, Q. Q. (2019). Genome-Wide Comparative Analyses of Polyadenylation Signals in Eukaryotes Suggest a Possible Origin of the AAUAAA Signal. *Int J Mol Sci, 20*(4). doi:10.3390/ijms20040958

Zhou, M., Palanca, A. M. S., & Law, J. A. (2018). Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat Genet, 50*(6), 865-873. doi:10.1038/s41588-018-0115-y

Acknowledgements

First and foremost, I would like to thank my mentor Selma Gago Zachert for providing the opportunity of pursuing my PhD. With her deep understanding of the topic and her fantastic mentoring, I gained the knowledge, skills, and confidence to excel in the field. I greatly appreciate the countless hours and effort she provided and will forever be grateful for the experience.

I would also like to thank my co-workers Susanne Engelmann and Shiv Meena for providing a friendly and relaxed work environment. Without our discussions and teamwork this work would have been substantially more difficult.

I want to thank the GRK 1591 for the providing mentoring and financial support. The input from top notch researchers and in addition the retreats provided an excellent environment to grow as a scientist. I especially want to thank Anne Baude for being for so friendly and for all the help over the years.

I thank the Leibniz Institute of Plant Biochemistry for providing an excellent research environment and tools.

Finally, without my family and friend none of this would have been possible. I thank my mother and father for always being loving and supportive. My brothers and sisters for advising me and being there through difficult times. I want to thank Chris, Nashrah, Nico, Mandi, and Jovy for all their support, I couldn’t have asked for better friends.

CURRICULUM VITAE

Name: Ammar Mousa Jaber
Date of Birth: May 10, 1988
Birthplace: San Francisco, California
Nationality: USA

EDUCATION

Leibniz Institute of Plant Biochemistry – Halle, Germany **Apr. 2015 – June 2019**
PhD Student

University of Maryland, Baltimore County – Baltimore, USA **Sept. 2011 – May 2014**
Masters in Professional Studies (MPS), Biotechnology (Cum laude)

Otto von Guericke Universität – Magdeburg, Germany **Oct. 2012 – Feb. 2013**
Study Abroad – Biosystemtechnik (Exchange Student)

University of Maryland, Baltimore County – Baltimore, USA **Sept. 2006 – May 2010**
Bachelor of Science (B.Sc.) in Biological Sciences

WORK EXPERIENCE

PAREXEL International – Baltimore, USA **July 2013 – Apr. 2015**
Research Associate

Fraunhofer IFF – Magdeburg, Germany **Feb. 2013 – June 2013**
Intern (Exchange Student)

University of Maryland, Baltimore County – Baltimore, USA **Apr. 2011 – Sept. 2011**
Research Assistant

AWARDS/FELLOWSHIPS

Graduate Fellow – German Research Foundation (GRK1519) **Apr. 2015 – June 2019**

Congress-Bundestag Youth Exchange for Young Professionals **July 2012 – July 2013**

OTHER SKILLS AND ACTIVITIES

Organizer of 13th annual Plant Science Student Conference – Halle, Germany **June 2019**
Member of Social Committee

Organizer of 13th annual Plant Science Student Conference – Halle, Germany **June 2017**
Head of Social Committee

Languages: English (Native), German (C1)

Statutory Declaration

Hiermit erkläre ich an Eides statt, dass ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe, die Arbeit selbständig und ohne fremde Hilfe verfasst, nur die angegebenen Quellen und Hilfsmittel genutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Halle (Saale), den _____ _____

Ammar Jaber