# Characterization of *cis*-natural antisense long noncoding RNAs overlapping the *UGT73C6* gene in *Arabidopsis thaliana*

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Doctor rerum naturalium)

der

Naturwissenschaftlich Fakultät I – Biowissenschaften – der

Martin-Luther-Universität

Halle-Wittenberg,

vorgelegt

von Herr

## Shiv Kumar Meena

geb. am 03.08.1986 in Bhawanpura, Rajasthan, India

PhD submitted on (Promotionsgesuch eingereicht am): 21.01.2020

Date of defense (Datum der öffentlichen Verteidigung): 25.08.2020

**Gutachter(s)**

1. Prof. Dr. Steffen Abel
   Department and Research Group, Leibniz Institute of Plant Biochemistry, Weinberg 3, Halle (Saale), 06120, Germany

2. Prof. Dr. Sven-Erik Behrens
   Institute of Biochemistry and Biotechnology, Faculty for Natural Science- I, Martin Luther University, Kurt-Mothes-Str. 3, 06120 Halle (Saale), Germany

3. Prof. Dr. Julia Kehr
   Faculty of Mathematics, Computer Science and Natural Sciences, Department of Biology, Institute for Plant Science and Microbiology
   Plant Molecular Genetics

*dedicated to my papa*

# Table of Contents

# Zusammenfassung

Long non-coding RNAs (lncRNAs) haben sich als wichtige Modulatoren der Genexpression in Eukaryonten erwiesen. In Pflanzen sind lncRNAs an einer Vielzahl biologischer Prozesse beteiligt, darunter Blütezeit- und Keimungsregulation, Wurzelentwicklung sowie bei der Hormon- und Stressantwort. Diese Forschungsarbeit konzentriert sich auf natural antisense long non-coding RNAs (NATlncRNAs), eine bestimmte Untergruppe von lncRNAs, die vom gegenüberliegenden DNA-Strang eines proteincodierenden Genes transkribiert werden. In der vorliegenden Arbeit charakterisierten wir zwei *NAT-lncRNAs* aus dem Modellorganismus *Arabidopsis thaliana*, bezeichnet als *lncNAT1*- und *lncNAT2*-, gemeinsam bezeichnet als *lncNATs-UGT73C6*, die das UDP-Glykosyltransferase-Gen *UGT73C6* überlappen. Es wurde bereits beschrieben, dass *UGT73C6* und sein nächstes Homolog *UGT73C5* eine Rolle in der Pflanzenentwicklung spielen, indem sie polyhydroxylierte steroidale Phytohormone, die Brassinosteroide (BRs) genannt werden, inaktivieren. Reporter-Genlinien, die die Promotorregion von *lncNAT1* und *lncNAT2* fusionieren, weisen auf eine unabhängige Promotoraktivität in Wurzeln und Sprossen hin. Die Analyse der *lncNATs-UGT73C6*-Transkripte zeigte, dass sie sehr stabil und cytosol-lokalisiert sind. Die Überexpression oder Runterregulierung jedes einzelnen *lncNATs-UGT73C6* beeinflusste die Rosettenblattfläche signifikant, während andere Entwicklungsprozesse, einschließlich Wurzellänge, Frischgewicht, Blattzahl und Samenertrag nicht beeinflusst wurden. Die beobachteten Phänotypen korrelieren jedoch nicht mit entsprechenden Veränderungen in den Transkriptlevel des überlappenden proteinkodierenden Gens *UGT73C6*. Zusätzlich reagiert *lncNATs-UGT73C6*, wie *UGT73C6* und *UGT73C5*, nicht auf die BR-Behandlung, aber Expressions- und in silico-Daten deuten darauf hin, dass sie ihre Funktion über eine Zielmimikry der microRNA396 ausüben können. Darüber hinaus konnten Peptide, die durch kleine offene Leserahmen in *lncNAT2* kodiert werden, nach transienten Expressionsassays in *Nicotiana benthamiana* nachgewiesen werden, obwohl ihre Überexpression keinen phänotypischen Effekt hat. Zusätzlich führt die Überexpression einer nicht-peptidcodierenden *lncNAT2*-Variante, bei der alle Startcodons mutiert waren, zu einer Vergrößerung der Blattfläche. Unsere Daten weisen darauf hin, dass *lncNATs-UGT73C6* als bona fide long non-coding RNAs wirken, die die Blattgröße bei *A. thaliana* modulieren.

# 1 Abstract

Long non-coding RNAs (lncRNAs) have been shown to be important modulators of gene expression in eukaryotes. In plants, lncRNAs are involved in a wide range of biological processes including flowering time and germination regulation, root development, hormone and stress related responses. This study focuses on natural antisense long non-coding RNAs (NAT-lncRNAs), a particular sub-type of lncRNAs that are transcribed from the opposite DNA strand of protein-coding genes. In the present work, we characterized two *NAT-lncRNAs* from the model organism *Arabidopsis thaliana*, referred as *lncNAT1* and *lncNAT2,* collectively *lncNATs-UGT73C6*, which overlap the UDP-glycosyltransferase gene *UGT73C6*. It has been previously described that *UGT73C6* and its closest homologue *UGT73C5* play a role in plant development by inactivating polyhydroxylated steroidal phytohormones called brassinosteroids (BRs). Reporter gene lines fusing the promoter region of *lncNAT1* and *lncNAT2* indicate independent promoter activity in roots and shoots, respectively. Analysis of *lncNATs-UGT73C6* transcripts showed that they are fairly stable and cytosol-localized. Overexpression or down-regulation of each *lncNATs-UGT73C6* significantly affected the rosette leaf area, whereas other developmental processes, including root length, fresh weight, leaf number, and seed yield were not affected. However, the observed phenotypes do not correlate with respective changes in transcript levels of the overlapping protein-coding gene *UGT73C6*. Additionally, *lncNATs-UGT73C6*, like *UGT73C6* and *UGT73C5*, remain unresponsive to BR treatment but expression and *in silico* data suggest that they can exert their function via target mimicry of *microRNA396*. Moreover, peptides encoded by small open reading frames present in *lncNAT2* could be detected after transient expression assays in *Nicotiana benthamiana* although its overexpression has no phenotypic effect. Additionally, the overexpression of a non-peptide coding *lncNAT2* variant, in which all the start codons were mutated, results in increased leaf area. Our data indicate that *lncNATs-UGT73C6* act as bona fide long noncoding RNAs modulating leaf size in *A. thaliana*.

## 2   Introduction

## 2.1   Long noncoding RNAs as modulators of gene expression

### 2.1.1   Noncoding RNAs

Until recently RNA molecules were considered as a mere intermediator for the flow of genetic information from gene to protein, a phenomenon generally referred to as the central dogma of molecular biology (Crick, 1958). Advent and subsequent progresses in next-generation sequencing technologies lead to genome-wide mapping and discoveries of different types of RNAs (Schadt et al., 2010) that challenged this dogmatic view. It is now a widely accepted fact that only a small proportion (Eddy, 2012) of total transcriptome is required for orchestrating the module for protein synthesis in the form of messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) in complex organisms such as mammals, mouse and plants (Okazaki et al., 2002; Carninci et al., 2005; Chekanova et al., 2007; Klepikova et al., 2016). Although protein centric view of molecular biology still holds its grounds, the debate that most of the genome undergo pervasive transcription in a developmentally controlled manner in virtually all organisms is setting the pace for the exploration of the roles of vast varieties of other unexplored noncoding RNA (ncRNA) species that do not undergo translation (Palazzo and Lee, 2015). Recent advancement in ncRNAs research has strongly contradicted the canonical 'junk RNA or transcriptional noise' conventions (Brosius, 2005) about the functionalities of ncRNAs (Mattick, 2009) due to our improved understanding that ncRNAs act as the major riboregulators. NcRNAs has been shown to modulate the expression of genes at multiple stages in numerous ways in a variety of organisms (Levine and Tjian, 2003; Henz et al., 2007). Independent of the advancement in ncRNA research, the narrative that RNA molecules were involved in prebiotic evolution and the origin of Life (Higgs and Lehman, 2015) has consistently fascinated the scientific community for persistent efforts to explore the evolutionary phenomenon such as the "RNA world" (Cech, 2012). NcRNAs comprise RNA molecules that do not encode proteins or small peptides and are categorized according to their size. Based upon the length of the transcript, whether shorter or larger than a defined cut-off of 200 nt, ncRNAs are classified into different sub-classes (Hombach and Kretz, 2016).

### 2.1.2 Classification of long noncoding RNAs

Examples of extensively studied small ncRNAs (i.e. <200 nt) include microRNA (miRNA), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), small nucleolar

RNAs (snoRNAs), and other short RNAs present in plants and animals (Kiss, 2002; Borges and Martienssen, 2015; Czech et al., 2018). In contrast, long noncoding RNAs (lncRNAs) are defined as transcripts longer than 200 nt, a size that excludes all known classes of small RNAs (Morris and Mattick, 2014), and does not encode for peptides longer than 70 amino acids (Ben Amor et al., 2009). Majority of lncRNAs contain bioinformatically predictable potential open reading frames (pORFs) that are usually not translated. Although, in some exceptional cases pORFs can undergo translation leading to the synthesis of peptides that are smaller than 100 amino acids in length (Nelson et al., 2016). LncRNAs are typically transcribed by RNA polymerase II (RNAPII) (Jensen et al., 2013), undergo splicing (Derrien et al., 2012) and possess canonical mRNA like features for RNA stability such as 5' cap and poly A tail at the 3' end of the transcript (Guttman et al., 2009). In addition to size-based criteria, regulatory ncRNAs are distinguished from housekeeping RNAs by categorizing cellular RNAs according to their function (Figure 1A). Besides, based upon the relative genomic location and nature of origin at transcriptional level, lncRNAs are further grouped into 4 different categories i.e. I) intronic lncRNAs are transcribed from the intronic region between two adjoining exons of a gene for e.g. *COLD ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*) (Heo and Sung, 2011) in *Arabidopsis* and *PROSTATE CANCER ASSOCIATED TRANSCRIPT-1* (*PCAT-1*) in humans (Prensner et al., 2011), II) intergenic lncRNAs (lincRNAs) are synthesized from a genomic region located between two adjacent genes e.g. *FLOWERING LONG INTERGENIC NON CODING RNA* (*FLINC*) (Severing et al., 2018) in *Arabidopsis* and 17kb long paradigmatic *XIST* (*X-INACTIVE SPECIFIC TRANSCRIPT*) lncRNA in mammals (Cerase et al., 2015; da Rocha and Heard, 2017). III) Promoter lncRNAs, which are transcribed in divergent manner from the promoter region of protein coding gene (Hamazaki et al., 2017; Severing et al., 2018) and IV) natural antisense (NAT) lncRNAs (NAT-lncRNAs) that are transcribed from the opposite DNA strand of a protein-coding gene thereby overlapping with cognate protein coding sense gene fully or at least partially in the exonic region (Figure 1 B). NAT-lncRNAs can acts in *cis* or *trans*. *cis* acting NAT-lncRNAs function by regulating the expression of complementary target gene or of the genes neighboring their site of transcription. For example *cis* acting *Tsix* (antisense transcript to *XIST*) downregulate the expression of *XIST* (Lee et al., 1999) that epigenetically induces silencing one of the X-chromosomes in vertebrate females for monoallelic expression and doses compensation (Cerase et al., 2015; da Rocha and Heard, 2017). In *Arabidopsis*, only a fewer example of functionally characterized *cis* acting NAT-lncRNAs are known. These examples include *COLD INDUCED LONG ANTISENSE INTRAGENIC RNA* (*COOLAIR*), *CYCLING DOF FACTOR 5* (*CDF5) LONG NONCODING RNA* (*FLORE*),

*ANTISENSE DELAY OF GERMINATION* (*asDOG1*) and *MADS AFFECTING FLOWERING4 ANTISENSE RNA* (*MAS*) (Swiezewski et al., 2009; Fedak et al., 2016; Henriques et al., 2017; Zhao et al., 2018).
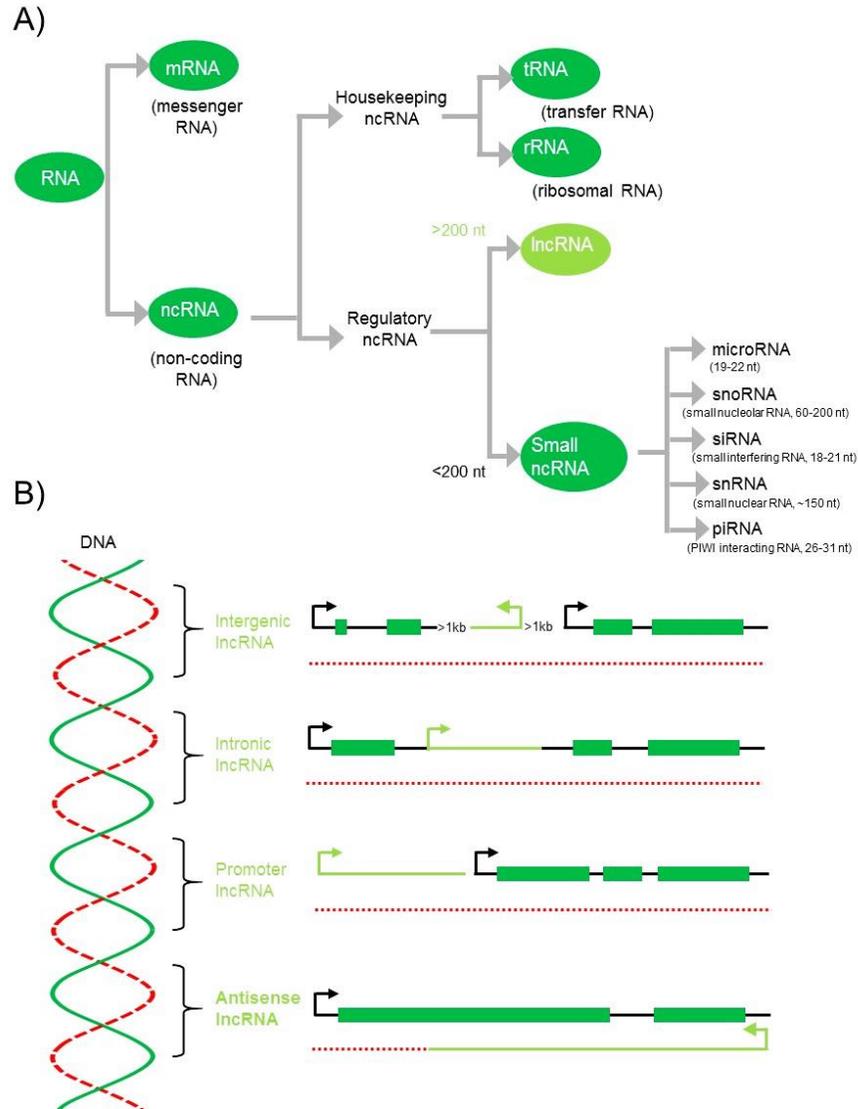


Figure 1: **Classification of RNAs**. A) Schematic for different types of RNAs. The size range of regulatory RNAs smaller than 200 nt is specified. RNAs termed as 'housekeeping' indicate their role in translation. B) Long noncoding RNAs (lncRNAs) classification based on their localization relative to nearby protein-coding genes and transcription origin. Intergenic lncRNAs (lincRNAs) are transcribed from locus situated between two genes and separated from them by at least 1 kb. Intronic lncRNAs initiate inside introns in either direction. Transcripts arising from the promoter region of a gene are termed as promoter lncRNAs. Antisense lncRNAs are transcribed from the opposite DNA strand of protein coding gene and overlap, partially or completely with them. Green boxes indicate exons. Black thin lines designate introns and 5' and 3' untranslated regions (UTRs) from protein coding genes. LncRNAs are shown as intron-less in fluorescent green. Arrowheads show 5' transcription initiation sites. Right handed unscaled 2 dimensional (2D) double helical twisted DNA strand is shown for structural impression of nucleic acid. Red dotted line(s) indicate other twin DNA strand in each symbolized gene type. nt: nucleotide(s), kb: kilo base. Part of Figure 1B is adapted after (Ariel et al. 2015).

While *trans* acting NAT-lncRNAs function independent of the site of synthesis and originate from a locus different than their overlapping partner gene for e.g. *BRAVEHEART* (Xue et al., 2016). *ANTISENSE HEAT SHOCK FACTOR HSFB2A* (*asHSFB2a*) in shown to act both in *cis* and *trans* over *HSFB2a* in *Arabidopsis* (Wunderlich et al., 2014). In contrast, it is not necessary that antisense strand will always encode for NAT-lncRNA. As exceptions, there also exist examples in animals and plants where both sense and antisense gene encodes for proteins (Terryn and Rouze, 2000; Su et al., 2012; Zhan and Lukens, 2013).

## 2.1.3 Functions and mechanisms of lncRNAs

A large number of lncRNAs have been annotated in different organisms. The lncRNAdb (http://lncrnadb.org) lists up-to date functionally validated lncRNAs (Quek et al., 2015). Until more than 1867 lncRNAs have been experimentally characterized where majority of them are from animal studies. It has been shown that most of lncRNAs are transcribed by RNA Pol II (Chekanova et al., 2007), also by Pol IV and V in plants (Wierzbicki et al., 2008). They undergo mRNA like post-transcriptional processing and soon after the biogenesis localize into the sub-cellular locations such as nucleus, cytoplasm and organelles. Apart from classification of lncRNAs based upon the nature of origin, it is possible to group lncRNAs according to localization and function (Chen, 2016). The localization of lncRNAs additionally has been suggested as an important predictor for the potential function in subcellular and physiological contexts. Evidences have shown that several nuclear retained lncRNAs can acts as chromatin remodelers via their association with chromatin modifiers, as splicing regulators or as epigenetic repressors or activators (Sun et al., 2018). Moreover, the act of transcription itself or the co-transcriptional processes such as splicing can be regulated by nuclear lncRNAs. Nucleus localized lncRNAs can influence the expression of target gene(s) via recruitment of various histone methyl transferases through their sequence guided interaction with polycomb repressive complex (PRC2) to deposit active heterochromatin H3K9 repressive mark at the promoter region of target gene. Moreover, they can act as activator of gene expression and thereby altering chromatin state due to histone 3 lysine 4 trimethylation (H3K4me) to activate or induce the expression of genes. In some cases lncRNAs themselves can form R-loop, a tri molecular association of DNA: RNA hybrid complex, to inactivate target gene(s) (Sun et al., 2013; Santos-Pereira and Aguilera, 2015; Marchese et al., 2017). In addition, the regulatory cascade of *cis* acting nuclear lncRNA can spread across the nearby genes to regulate their expression via recruitment of protein or by following other unknown mechanisms. Moreover lncRNAs can play an important role in the organization of nuclear domains (Sun et al., 2018). Other mechanisms by which lncRNAs can function is acting as 'decoys' to

modulate gene expression by titrating out chromatin remodeling complexes or histone methyl transferases or deacetylases from the physical association with the target gene loci (Fan et al., 2015; Jain et al., 2016).

On the other hand, cytoplasmic lncRNAs can adopt a variety of means to regulate expression of genes post-transcriptionally. Indeed several signaling cascades can be directly or indirectly regulated by lncRNAs via mRNA degradation, sponging of ribonucleoproteins and turnover of proteins (Wang and Chekanova, 2017a). Also, many cytoplasmic lncRNA can act as decoys for microRNAs (miRNA) activity (Franco-Zorrilla et al., 2007). The translational efficiency of target mRNA have also been shown to be enhanced by cytoplasmic lncRNAs (Jabnoune et al., 2013). Since a large number of lncRNAs achieve their function via interaction with proteins, the association of lncRNA with RNA binding proteins (RBP) can allow their translocation within regulatory cascades to participate in gene expression regulatory networks (Wang and Chekanova, 2017a). A summary of various mechanisms for cytosolic lncRNAs is outlined in Figure 2.

Thousands of lncRNAs are annotated in plants. However, in comparison to animal studies, only a less than two dozen of plants lncRNAs are functionally characterized including lncRNAs present in the model plant *A. thaliana* and crop models such as rice (*Oryza sativa*) and tomato (*Solanum lycopersicum*). Examples of studied lncRNAs in plants shows their role in various biological processes for e.g. in root development, response to light, flowering time regulation, reproduction, and stress responses (Ben Amor et al., 2009; Ietswaart et al., 2012; Liu et al., 2012; Di et al., 2014; Li et al., 2014; Wang et al., 2014; Zhang et al., 2014; Li et al., 2016; Yuan et al., 2016). Notwithstanding, despite being one of the most abundant lncRNAs class, only a limited sets of NAT-lncRNAs have been functionally characterized in plants. Though more than 30000 NAT-lncRNAs pairs are estimated in *A. thaliana* (Wang et al., 2014), the number of experimentally validated NAT-lncRNAs still remains abysmal in spite of their discovery as early as in 1976 first in viruses (Barrell et al., 1976) than in prokaryotes (Tomizawa et al., 1981) and eukaryotes (Williams and Fried, 1986) during the nineties. The best-studied examples include the lncRNAs that are involved in flowering time regulation. NAT-lncRNAs *COOLAIR* and intronic lncRNAs *COLDAIR* are expressed in *A. thaliana* during vernalization, and after a prolonged exposure to cold, function in synergistic manner to epigenetically silence *FLOWERING LOCUS C* (*FLC*). Intronic *COLDAIR* recruits PRC2 to *FLC* locus via its physical interaction with polycomb group protein EZH2 methyltransferase subunit of PRC2 complex and deposits repressive chromatin marks (Heo and Sung, 2011). The expression of *COOLAIR,* a pool of alternatively spliced and differentially polyadenylated antisense transcripts synthesized from the opposite DNA strand of

*FLC,* peaks during cold and inhibits *FLC* expression by deposition of repressive H3K27me3 marks. This process facilitates switching of chromatin states and allows flowering in *A. thaliana* (Swiezewski et al., 2009; Csorba et al., 2014). Moreover, expression of antisense *COOLAIR* and sense *FLC* is mutually exclusive thereby resulting in the expression of either *COOLAIR* or *FLC* from one allele (Rosa et al., 2016). Additionally, *COLD OF WINTER-INDUCED NONCODING RNA* (*COLDWRAP*) that is transcribed from the upstream region of *FLC* promoter has also been described to suppress *FLC* (Kim and Sung, 2017). *MAS,* is another NAT-lncRNA that has been reported to play a role in preventing precocious flowering. *MAS* is transcribed from opposite DNA strand of *MAF4* locus and recruits a protein subunit of COMPASS-like complexes, WDR5a, to *MAF4* locus to enrich H3K4me3 histone marks indicating its transcriptional activation (Zhao et al., 2018).

Following section outlines other reported NAT-lncRNAs, in *A. thaliana* and other plant species, that have been shown to control crucial developmental processes, but the mechanisms are still under study. For example *FLORE* NAT-lncRNA expression in *A. thaliana* oscillates with circadian rhythms and modulates expression of the sense gene *CDF5*. The opposing expression profiles of *FLORE* and *CDF5* forms a negative feedback loop and thus helps prevent inhibitory effects of *CDF5* to progressively switch plant for flowering stage by allowing the expression of *FLOWERING LOCUS T* (*FT*). Also, the suppressive role of *FLORE* in the downregulation of other *CDFs* viz: *CDF1*, *CDF3* and *CDF5* suggest that its regulatory cascade spread over the locus and not limited to the sense gene. The molecular mechanism of mutual inhibitory relationship between *FLORE* and *CDFs* still needs to be identified. It was clearly shown that *FLORE* effects are not mediated by generation of small-interfering RNA (siRNA) (Henriques et al., 2017), a mechanism that was initially proposed as main means of NAT-lncRNAs action. Another descriptive study showed that *asDOG1* is derived from complementary *DOG1* strand and consequentially act as negative regulator of seed dormancy in *Arabidopsis* by inhibiting expression of *DOG1* in *cis*. However, overexpression studies showed that *asDOG1* is unable to produce similar effect in *trans* (Fedak et al., 2016). Similarly, heat induced *asHSFB2a* and its overlapping protein coding gene *HSFB2a,* a heat stress response factor, are antagonistically expressed from the locus to regulate the development of female gametophytes in *Arabidopsis*. Ectopic overexpression of either of *asHSFB2a* and *HSFB2a* counters expression of each other in a manner similar to Yin-Yang model of gene expression regulation. Nonetheless, the mechanism of gene expression regulation by *asHSFB2a* is not yet known (Wunderlich et al., 2014).
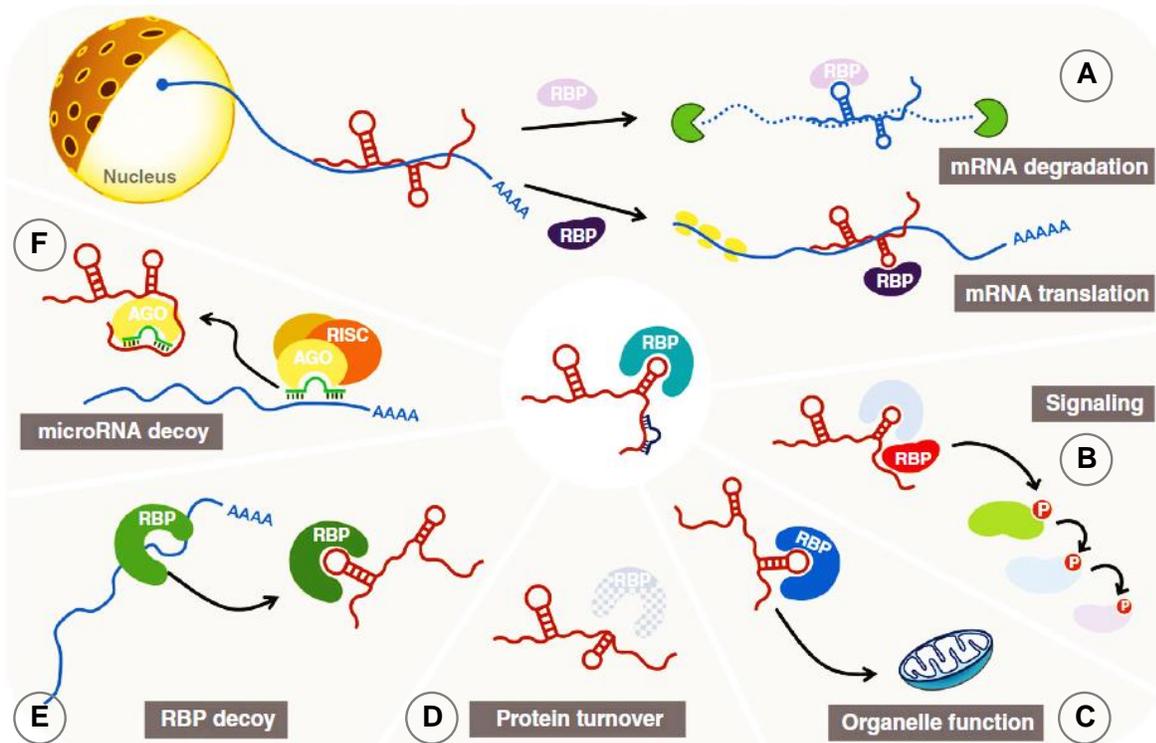
Figure 2: **Cartoon representation showing different layers of gene expression regulation by cytoplasmic lncRNAs**. (Clockwise) Cytoplasmic lncRNAs are localized to cytoplasm after transcription and can acts via interaction with RNA-binding proteins (RBPs) or with partially complementary mRNAs to influence stability and/or translation of target mRNAs (A). Activation of particular signaling cascades can be achieved by association of lncRNAs with RBPs that activates kinases (B). Moreover, lncRNA-RBP complex can regulate organelle functions (C). Also, lncRNAs can serve as platform for RBPs to regulate protein turnover (D). Furthermore, lncRNAs can act as decoy for RBPs (E) or target mimic microRNAs (F) thereby regulating functional output. (Figure reproduced after Noh et al., WIREs RNA, 2018).

Apart from developmental role of lncRNAs, a few of them have been described to be involved in stress and defense related responses in plants. *ELENA, ELF18-INDUCED LONG-NONCODING RNA,* induces expression of *PATHOGENESIS-RELATED GENE1* (*PR1*) to increase pathogen resistance via its interaction with *Arabidopsis MEDIATOR COMPLEX SUBUNIT 19A* (*MED19a*) thereby titrates out its negative regulator *FIBRILLARIN 2* (*FIB2*) from *FIB2/MED19a* complex at *PR1* promoter (Seo et al., 2017; Seo et al., 2019). MED19a is well known to arbitrate interactions between transcriptional activators and RNA Pol II in defense responses (Wang et al., 2008; Kidd et al., 2009; Mathur et al., 2011). In tomato as well, *NAT-lncRNA 16397* has been shown to render resistance against late blight disease by inducing positive regulation of sense target gene *SlRX22,* a glutaredoxin gene family member, which results in reduced infection by *Phytophthora infestans* (Cui et al., 2017). Several other NAT-

lncRNA are predicted and shown to be induced by other pathogens such as *Fusarium oxysporum* in *Arabidopsis* (Zhu et al., 2014). In rice (*Oryza sativa)* NAT-lncRNAs control developmental and stress responses. For example *TWISTED LEAF* regulates lead blade flattening via downregulation of overlapping sense gene *OsMYB60*, member of conserved gene family of transcription factors that play role in development and stress responses (Dubos et al., 2010), supposedly by chromatin modification at *OsMYB* locus (Liu et al., 2018). Also in rice, phosphate starvation (Pi) induced *cis-NAT*$_{PHO1;2}$ has been shown to positively affect *PHOSPHATE$_{1;2}$* (*PHO$_{1;2}$*) mRNA translation by associating with ribosomes without affecting steady state mRNA levels of *PHO$_{1;2}$*. This process facilitates formation of a feed-forward loop for increased loading of Pi in xylem vasculature against Pi deficiency (Jabnoune et al., 2013). Indeed, using polysomal profiling approach approximately 14 *cis*-NATs and more than 100 *trans*-NATs have been predicted in *A. thaliana* that could be involved in positive or negative translational regulation of various target genes (Deforges et al., 2019b; Deforges et al., 2019a). One of the other widely anticipated mechanism of NAT-lncRNAs is based on the generation of natural antisense siRNAs (nat-siRNAs) (Borsani et al., 2005; Held et al., 2008) due to formation double stranded RNA (dsRNA) between NAT-lncRNA and complementary target transcript. Subsequent slicing activity by DICER complex produces 21-25 nt siRNAs that can post-transcriptionally inhibit expression of sense gene. However, new studies highlight that this mechanisms for NAT-lncRNAs is an exception rather than a rule (Ariel et al., 2015). Studies suggest role of NAT-lncRNAs in epigenetic modifications independent of small RNA pathway (Luo et al., 2013). Unlike predictable RNA-RNA hybridization, NAT-lncRNAs can achieve their function via RNA-protein interactions (Mattick, 2005; Willingham et al., 2005). As mentioned before, lncRNAs in cytoplasm can act as decoys to competitively inhibit actions of ribonucleoproteins (RBPs) and miRNAs. A variety of signaling cascades can be directly or indirectly regulated by lncRNAs (Wang and Chekanova, 2017a). Nevertheless, lncRNAs mechanism are far diverse than originally speculated and accumulating examples are adding to the already existing complex layers of gene expression regulation.

## 2.2   *cis-natural antisense transcripts* of *UGT73C6* in *A. thaliana*

### 2.2.1   *UGT73C* subfamily in *UGT* multigene family

Uridine diphosphate (UDP) glycosyltransferases (UGT) are carbohydrate-active enzymes (CAZy) that are ubiquitously present in animals and plants. UGTs catalyze attachment of an UDP-activated donar glucose moiety to various aglycone substrates, a process termed glycosylation. Glycosylation of aglycon substrates such as plant hormones, secondary metabolites, biotic and

abiotic toxic substances etc. has been shown to alter the bioactivity, solubility and other physical properties associated with the storage and transport of endogenous metabolites in cellular environment (Ross et al., 2001; Poppenberger et al., 2003; Lim and Bowles, 2004; Husar et al., 2011; Li et al., 2018; Haroth et al., 2019). Therefore, glycosylation is one of the adaptive means to maintain homeostasis in plants (Li et al., 2001). The enzymatic reaction of UGTs is mediated by the presence of a conserved carboxy-terminal signature sequence known as plant secondary product glycosyltransferase (PSPG) motif (Paquette et al., 2003). Key amino acid residues in PSPG motif forms a hydrophobic pocket and expedite binding with the donar sugar molecule followed by formation of a region-specific acceptor pocket to glycosylate aglycones in the active site. Different amino acid residues across the polypeptide chain in UGTs contribute to intricate 3D-dimensional structure for the formation of the acceptor pocket and the active site. Binding of aglycones results in conformational changes leading to the attack of unprotected nucleophilic hydroxyl group from acceptor molecule over the donar carbon of the oxocarbenium ion formed during the reaction. The catalytic cascade finally leads to formation of position specific O-glycosidic bond between the sugar moiety and glycosylated target (Osmani et al., 2008). The acceptor pocket formation is region specific and recognizes only the overall backbone of target substrate rather than molecule itself specifically. As a result, UGTs in general tend to act promiscuously *in vivo* and does not exhibit peculiar substrate specificity (Vogt and Jones, 2000; Richman et al., 2005; Osmani et al., 2008).

## 2.2.2 Role of *UGT73C5* and *UGT73C6* in the regulation of BR homeostasis

The genome of *A. thaliana* possess about 119 UGTs genes that are classified into 14 distinct subgroups on the basis of their sequence similarity and presence of conserved PSPG motif (Ross et al., 2001; Langlois-Meurinne et al., 2005). The *UGT73C* subfamily belongs to group D of the UGT superfamily and consists of seven closely related genes. *UGT73C1, UGT73C2, UGT73C3, UGT73C4, UGT73C5* and *UGT73C6* are clustered in tandem repeats on chromosome 2 while *UGT73C7* is located on chromosome 3 (Figure 3). All seven subfamily members are devoid of introns and their nucleotide sequences shows a range of 77.6-91% identity. Due to this high level of similarity it has been strongly suggested that they are tandem duplicates, a major gene duplication mechanism in eukaryotes (Reams and Roth, 2015). *UGT73C7* might be a result of a gene duplication event (Ross et al., 2001). Therefore, these genes might share analogous regulatory elements and are speculated to have similar functions.
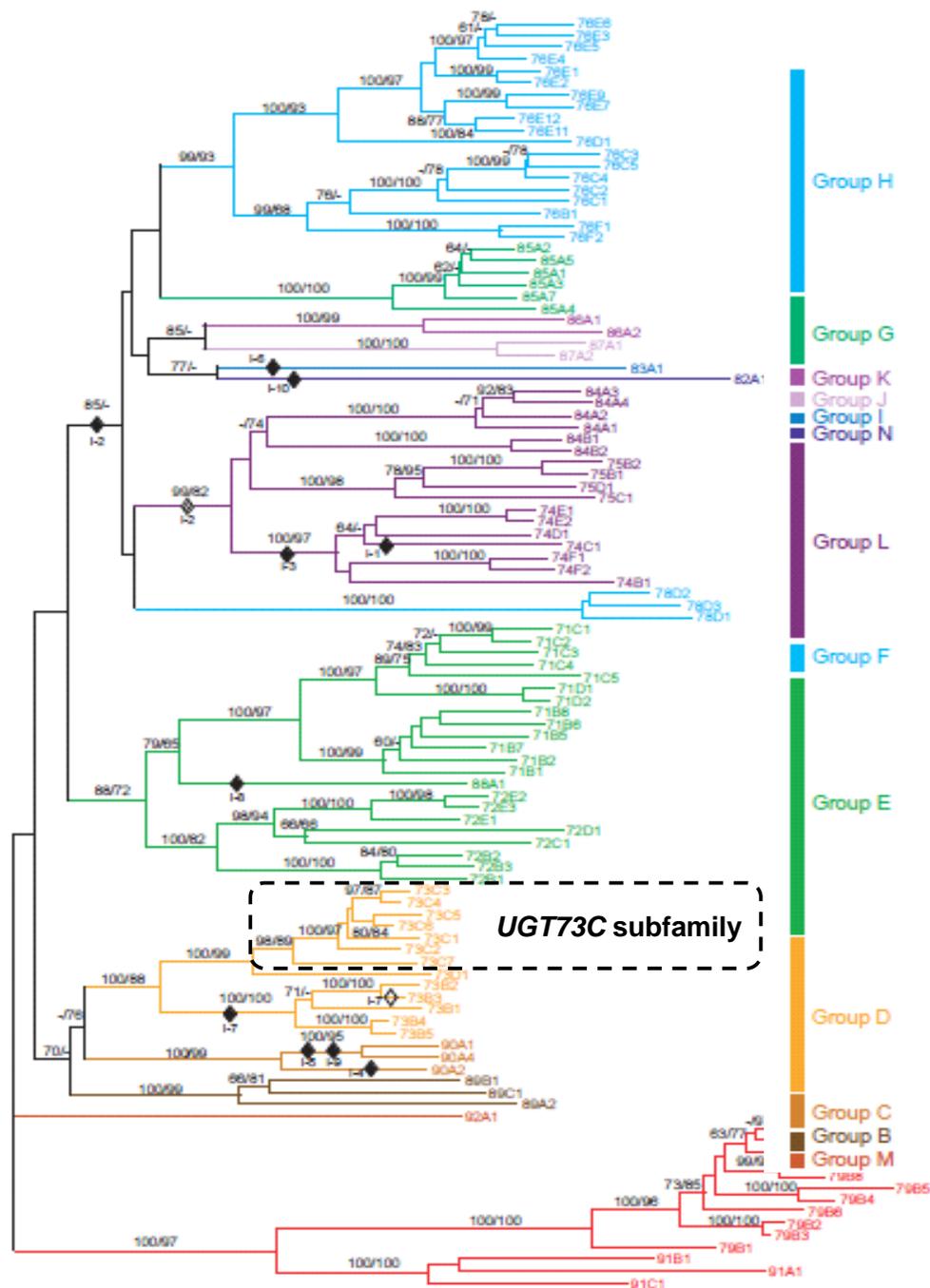
Figure 3: **Phylogenetic tree of UGT family-1 of *Arabidopsis thaliana*.** UGT family contains more than 119 genes sub-grouped in various clades (A to N) based upon sequence similarity. Each UGT is characterized by the presence of a conserved PSPG motif. The *UGT73C* subfamily contains 7 genes (highlighted in black rectangle edge squire) and *UGT73C6* is overlapped by two natural antisense long noncoding RNAs (NAT-lncRNAs) viz: *lncNAT1* and *lncNAT2*, collectively referred as *lncNATs-UGT73C6*. Figure is modified after (Ross et al., 2001).

*In vitro* characterization of *UGT73C* subfamily member's activity showed that they can recognize more than one aglycon substrate. The analysis of catalytic activities showed broader substrate specificity for *UGT73C* subfamily members. Various secondary metabolites, plant hormones, fungal mycotoxins and xenobiotics are recognized as potential aglycon substrates. For example UGT73C1 and UGT73C5 are able to glycosylate class of key phytohormone called cytokinins (CKs) under *in vitro* conditions and were shown to form *O*-glycososides with *trans*-zeatin, *cis*-zeatin and dihydrozeatin (Hou et al., 2004). As a result, glycosylated CKs turns inactive and serves as stable form of storage in cellular environment for this important growth promoting plant hormone (Hou et al., 2004). However, *in vitro* studies can only provide a preliminary read out for potential catalytic activities *in planta* (Bowles et al., 2005). Previous attempts have been made to characterize *UGT73C1, UGT73C2, UGT73C3, UGT73C4, UGT73C5* and *UGT73C6 in planta* by means of ectopic overexpression. Unlike other sub-family members, only plants overexpressing *UGT73C5* or *UGT73C6* were able to exhibit obvious morphological phenotypic effects (Husar et al., 2011). Overexpression of *UGT73C5* or *UGT73C6* showed a typical brassinosteroid (BR) deficiency phenotype marked by the presence of dark green and cabbage leaf morphology (Azpiroz et al., 1998; Poppenberger et al., 2005; Husar et al., 2011). Husar *et. al*. emphasized in their analysis that *in vivo* enzymatic activities and phenotypic effects of *UGT73C5* and *UGT73C6* are akin to each other. *UGT73C5* and *UGT73C6* are 91% similar and their encoded proteins are shown to convert biologically active forms of brassinosteroids (BRs) viz: casterone (CS) and brassinolide (BL) into glycosylated ones by means of 23-*O*-glycosylation in exogenous feeding experiments with CS and BL. This resulted in higher levels of BL-23-O-glycoside and CS-23-O-glycoside in plants overexpressing *UGT73C5* or *UGT73C6* compared to the wild type (WT) controls. Moreover, BRs-23-*O*-glycoside were found to additionally form BRs-malonylglucosides suggesting a potentially different physiological role for *UGT73C6* and *UGT73C5 in planta*. (Poppenberger et al., 2003; Husar et al., 2011).The glycosylated BRs potentially represent storage form in cell and, therefore, no longer participate in the associated signaling processes related to cell division, elongation, and differentiation which consequently hampers the normal growth and development of plants thereby resulting in characteristic cabbage phenotype (Clouse et al., 1996). Thus, these finding highlighted that *UGT73C5* and *UGT73C6* participates in regulation of BR homoeostasis *in planta* and that both UGT73C5 and UGT73C6 inactivate most active forms of BRs by glycosylation (Poppenberger et al., 2005; Husar et al., 2011).

## 2.2.3 *Antisense long noncoding RNAs* of *UGT73C6*

As many as more than ten thousands of NAT-lncRNAs are predicted in *A. thaliana* (Wang et al., 2014). The UGT superfamily contains 7 annotated sense-antisense pairs spread across different phylogenetic clades. The relevance of these *cis*-NAT-lncRNAs pairs in the multigene family context has not yet been elucidated. Two natural antisense transcripts (source: Araport11) viz: *NAT1-UGT73C6* (*lncNAT1*) and *NAT2-UGT73C6 (lncNAT2),* collectively referred as *lncNATs -UGT73C6* in this work, are encoded by the complementary DNA strand of the *UGT73C6* gene (1677 bp) Figure 3 and 4. Based on TAIR 10, lnc*NAT1* transcript is 505 nt longer than lnc*NAT2* (1084 nt) owing to differences in transcription start sites. Both longer and shorter variants exist in spliced and unsliced forms and contains approximately 98 bp and 73 bp long intron respectively. Annotated *lncNAT1* shares an overlapping region of 1512 bp while *lncNAT2* shares 986 bp sequence overlap with *UGT73C6*. In addition, *lncNATs-UGT73C6* shares high sequence complementarity with other *UGT73C* sub family members in clade D. *UGT73C5* is ~90% complementary to *lncNATs-UGT73C6* while sequence similarity from *UGT73C1* to *UGT73C4* is around 80 %. *UGT73C7,* which is located on chromosome 3, has 68% complementary to *lncNATs-UGT73C6*. Moreover, *lncNAT1* overlaps and shares high levels of sequence complementarity with the PSPG motif of all the *UGT73C* family members. The extent of complementarity between *lncNATs-UGT73C6* and various *UGT73C* subfamily members in listed in Table 1. High levels of sequence complementarity between sense and antisense pairs have been previously speculated to be responsible of the formation of gene regulatory circuit between NATs-lncRNAs and closely related genes (Wang et al., 2006). In fact, dsRNA resultant from the co-expression of NATs-lncRNA and cognate protein coding genes, can be recognized as a substrate by DICER complex leading to the generation of 21-24 nt long nat-siRNAs. nat-siRNAs can be further loaded onto the effector RNA-induced silencing complex (RISC) and could direct the AGO-mediated cleavage of complementary targets genes similar to earlier reports (Borsani et al., 2005; Held et al., 2008). Therefore, *lncNATs-UGT73C6* could establish a local gene expression regulatory loop to modulate the gene expression of not only *UGT73C6* but also other closely related *UGT73C* family members via nat-siRNAs. Due to sequence complementarity, nat-siRNAs mediated silencing mechanism was hypothesized as one of the potent and most widespread means of action for several thousands of NATs-lncRNA not only in *A. thaliana* but also in other species based upon previous reports (Borsani et al., 2005; Held et al., 2008). However, as stated earlier also, several reports suggest that NATs-lncRNA can alter target gene expression via mechanisms other than the anticipated siRNA pathway for e.g. epigenetic modifications and modulation of translational

efficiency of protein coding gene (Wang and Chekanova, 2017b). Nonetheless, the number of functionally corroborated NAT-lncRNAs or lncRNAs in general remains very low and, thus, *de novo* characterization of any new lncRNA is relevant to further enhance our understanding of the functions and mechanisms of lncRNAs.
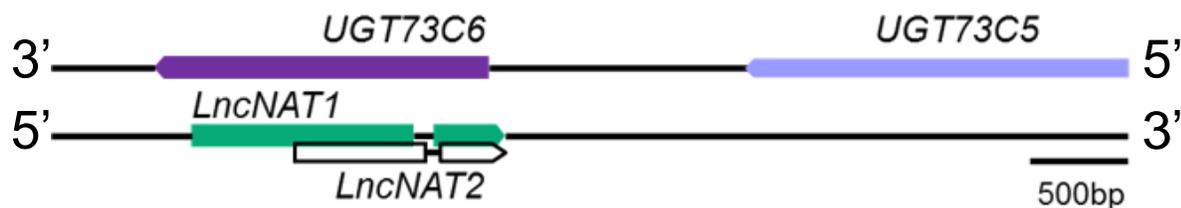


Figure 4: **Schematic representation of genomic locus showing *UGT73C6* (violet), *lncNAT1* (green), *lncNAT2* (white) and closest homologue *UGT73C5* (light blue).** Introns and intergenic region is shown as thick solid black line. For simplified illustration, 5' and 3' untranslated regions in *UGT73C6* and *UGT73C5* are not shown.

## 2.3  Perspectives on leaf morphogenesis in *Arabidopsis*

LncRNAs can play miscellaneous roles in a number of molecular networks controlling responses to development, environmental cues, stresses, biotic and abiotic factors. In the context of investigations carried out in the presented thesis work, role of *lncNATs-UGT73C6* was examined from several angles. A phenotypic analysis was carried out during the progression of various developmental stages throughout the life cycle of *A. thaliana*. Alterations in the levels of *lncNATs- UGT73C6* resulted in modulation of rosette area (Figure 9 and 10). Thus, in order to facilitate a better comprehension of the phenotypic effects of *lncNATs-UGT73C6*, a brief overview of the processes and factors that govern leaf development is summarized in subsequent sections.

Leaves are the primary organs in seed plants that carry out photosynthesis, a process of conversion of carbon dioxide ($CO_2$) and water ($H_2O$) into the organic sugars (glucose) and molecular oxygen ($O_2$) (Ingenhousz, 1779; Johnson, 2016). Photosynthesis not only forms a basis of sustaining the oxygen content in earth's atmosphere but also act as a supplier for most of the energy demand in the form of organic compounds for existence of life on Earth (Bryant and Frigaard, 2006). In the absence of leaves, plants cannot harvest light energy, synthesize organic compounds in addition to the inability to respond to biotic and abiotic stress factors such as pathogens and light quality and quantity. Moreover, flowers are modified leaves (Goethe, 1790; Pelaz et al., 2001) and therefore plants will be unable to perpetuate without leaves. Study of leaf development has been of great interest for many investigators and attempts to genetically dissect

morphogenesis of leaves has been made in past decades mostly using eudicot model *A. thaliana*. Though a complete understanding of leaf development still at parse, nonetheless recent breakthroughs using both molecular and genetics tools have uncovered previously unexplored complex gene regulatory aspect of leaf development in extensive details in *Arabidopsis* and other species (Tsukaya, 2002b, 2013b; Kalve et al., 2014; Du et al., 2018).

Though leaves are apparently simple and flat structures, their tightly controlled morphogenesis is complex and involves a series of coordinated interplay of gene networks during different stages of development. Conceptually early leaf development consists of four synchronized and possibly overlapping growth phases in eudicots viz: **a**) initiation of leaf primordium from shoot apical meristem (SAM), **b**) distal growth after initiation of leaf primordium and establishment of dorso-ventral polarity i.e. formation of adaxial-abaxial and proximal-distal axes, **c**) growth of leaf blade, also known as lamina, along the medio-lateral axis, and **d**) multidirectional intercalary growth of lamina that expands leaf dimensions (Foster, 1936; Poethig and Sussex, 1985; Steeves and Sussex, 1989; Smith and Hake, 1992; Donnelly et al., 1999; Ichihashi et al., 2011; Nakata and Okada, 2013). Incipient primordium or founder cell, and thus all the lateral organs, originates from the peripheral zone (PZ) of SAM that consists of PZ and an inner cortical zone (CZ). Expression of *SHOOT MERISTEMLESS* (*STM*) encoded class I KNOTTED-LIKE HOMEOBOX (KNOX1) transcription factors (STM: KANT1, KNAT2 and KNAT6) are key to the formation of SAM and maintenance of pluripotent fate of cells in CZ (Long et al., 1996). The maintenance of stem cell identity and meristematic activity in CZ of SAM itself is also facilitated by expression of *WUSCHEL-RELATED HOMEOBOX* (*WOX*) genes. *WUSCHEL* (*WUS*) is founding member of *WOX* gene family that are indispensable to meristematic zones in plant body and their upstream regulation in apical (tip), intercalary (middle), and lateral (sides) meristems is dependent over the peptide ligands viz : CLAVATA (CLV1), CLV2 and CLV3. An unknown signal activates *WUS* expression of *CLV3,* also known as *ENDOSPERM SURROUNDING REGION* (*CLE*), which in turn leads to further binding with CLV1/2 thereby resulting in inhibition of *WUS*. Other reports also suggest that CLE family peptides interacts with leucine-rich receptor-like kinases (LPR-RLKs) to achieve regulation of *WOX* family members (Katsir et al., 2011). These regulatory components play crucial role for switching of stem cell fate into the actively dividing fate. A periodic auxin maximum is established in the flanking regions of SAM, i.e. PZ, by auxin efflux carrier PINFORMED1 (PIN1). Due to PIN1 activity PZ cell fate is destined to develop into leaves. Concurrent asymmetric growth and extensive cell proliferation in PZ than results in leaf protrusion. Though *KNOX1* is required for SAM cell fate (Long et al., 1996),

its expression is repressed and sustained in later developmental stages, as the leaf grows, by repressive multiprotein complex of ASYMETRIC LEAF1 (AS1) and AS2 (Xu et al., 2003), a LATERAL ORGAN BOUNDARIES (LOB) domain protein. AS1-AS2 complex physically interacts with the promoters of *KNOX1* genes viz: *BREVIPEDICELLUS* (*BP*) and *KNAT2* (Guo et al., 2008). Once the leaf blade and petiole identity is established, the leaf margins development takes place. Boundary region is formed due to the reduced rate of cell division and growth. KNOX1 proteins play positive role in boundary formation. *mir164* regulated CUP-SHAPED COTYLEDON (CUC) NAC domain transcription factors CUC1, CUC2 and CUC3 affects border formation. *miR164* promotes cleavage of *CUC1* and *CUC2* mRNA. As a result, a regulatory feedback loop comprising *mir164* and *CUC* is formed during boundary morphogenesis (Vroemen et al., 2003; Laufs et al., 2004; Mallory et al., 2004; Hibara et al., 2006). Subsequent to the boundary formation, molecular players critical for simultaneous adaxial-abaxial (Ad-Ab) polarity have been characterized genetically. Processes of Ad-Ab patterning and emergence of leaf primordium occurs concurrently. Classical surgical experiments performed in 1951 by Sussex showed that a peripheral microincision of leaf primordium leads to the emergence of alternative cylindrical and abaxialized primordium (without a leaf blade) suggesting that a SAM signal, termed as 'Sussex signal', is indispensable to Ad-Ab leaf polarity (Sussex, 1951). However studies involving amputation of incipient leaf from SAM in *A. thaliana* or lateral incision in the region flanking leaf primordium in other species contradicts origin of Sussex signal exclusively from SAM. Similar polarity defects occur in the form of abaxialized radially symmetric leaf (Reinhardt et al., 2005) because incision in flanking regions does not preclude communication of primordium with SAM (SNOW and SNOW, 1959; Shi et al., 2017). However, despite enormous efforts, the molecular identity of Sussex signal is yet to be specified (Du et al., 2018). Domain specific expression and mutual repression of Ad-Ab genes by distinct classes of transcription factors and small RNAs is considered crucial for subsequent maintenance and reinforcement of Ad-Ab polarity. These genes includes class III Homeodomain-Zinc finger (HD-ZipIII) family *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*) and *REVOLUTA* (*REV*) in Ab side. Ad identity is regulated by *KANADI* (*KAN*) family (*KAN1* and *KAN2*) and *ETTIN* (*ETT*)/*AUXIN RESPONSE TRANSCRIPTION FACTOR* (*ARF*) *3* and *ARF4*. Furthermore, tasiR-ARF, which are generated by the cleavage activity of *miR390* over *TRANS ACTING CIS RNA3* (*TAS3*) mRNA, induce the degradation *ARF3* and *ARF4* transcripts. A brief outline of the processes involved in leaf development are superficially outlined in Figure 5.

At the beginning of leaf protrusion, Ad-Ab identity establishment and intercalary lamina growth is mostly accompanied by an enhanced cell proliferation that continues throughout leaf

morphogenesis. Several gene are described to be involved in the control of cell proliferation. The Arabidopsis *GROWTH REGULATORY FACTORS* (*AtGRFs*)/*GRF-INTREACTING FACTOR1*/*ANGUSTIFOLIA* (*AN3*) and *mir396* module broadly acts during the entire leaf growth. *GRFs-GIFs* (*GIF1*, *2* and *3*) delay transition from cell proliferation to cell differentiation by affecting levels of *CYCLINB1;1*, *CYCLIND3;1* and *KNOLLE* (Rodriguez et al., 2010; Debernardi et al., 2014). AN3/GIF1 coactivator, which is expressed in mesophyll cells and moves to epidermis cell layers as a protein (Kawade et al., 2013), synergistically interacts with chromatin remodelers to facilitate spatiotemporal GRF-GIF activities, and *mir396* regulate abundance of *GRFs* post-transcriptionally. *miR396* promoted cleavage activity of *GRFs* mRNA results in two opposing gradient of *miR396* and GRFs from tip to bottom of the leaf (Kim et al., 2003; Jones-Rhoades and Bartel, 2004; Kim and Kende, 2004; Horiguchi et al., 2005; Debernardi et al., 2012; Debernardi et al., 2014). Constitutive overexpression of *mir396* or *miR396* cleavage resistant *GRFs* decreases and increases leaf area respectively (Kim et al., 2003; Kim and Lee, 2006; Rodriguez et al., 2010). In contrast, class II TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) are another set of transcription factors that play crucial role by repressing meristematic activity and promoting transition from cell proliferation to cell expansion. *TCP4* acts upstream of *GIF1* and *GRFs* and represses genes that are not targeted by *miR396* including numerous *CUC* genes (Koyama et al., 2017). On one hand expression of *TCPs* itself is post- transcriptionally downregulated by *miR319* (Palatnik et al., 2003) while on other hand *TCP4* directly upregulates *mir396* (Rodriguez et al., 2010; Schommer et al., 2014). Thus, a balanced marginal and complete leaf growth is achieved via both *mir396*-*GRF* and *mir319*-*TCP* gene networks.

Apart from above mentioned gene networks, several other important gene modules participate in the tightly regulated intrinsic development programs for positive and negative regulation of leaf primordia growth via their action over meristematic activities, cell proliferation and differentiation. An initial voluminous enlargement in the cell dimensions relies largely upon the cytoplasmic growth while cellular growth during later stages includes substantial increase due to vacuolar growth. Environmental cues such as light and changes in metabolic states for sugars trigger biosynthesis of auxin in shoot apex thereby associated above mentioned *PIN1* mediated developmental patterning during early growth. These signals merge in SAM for lateral organ formation and converge on a central growth regulator, TARGET OF RAPAMYCIN (TOR) i.e. a Serine/Threonine kinase of the phosphatidylinositol-3-kinase-related kinase (PIKK) kinase family. TOR is involved in a vast range of cellular responses such as integration of central metabolic pathways, glucose signaling, biogenesis of ribosomes, initiation of translation, cell proliferation,

differentiation and autophagocytosis during the cytoplasmic growth (Zhang et al., 2013). Unlike in mammals and yeast, TOR pathway is not well understood Plants. Studies from poorly characterized TOR signaling in *Arabidopsis* suggest a key role of *TOR* in cell division, cell cycle progression and, thus, in active cell growth and light activated auxin/PIN1 mediated pattern generation for organ growth (Moreau et al., 2012; Xiong and Sheen, 2012; Li et al., 2017).
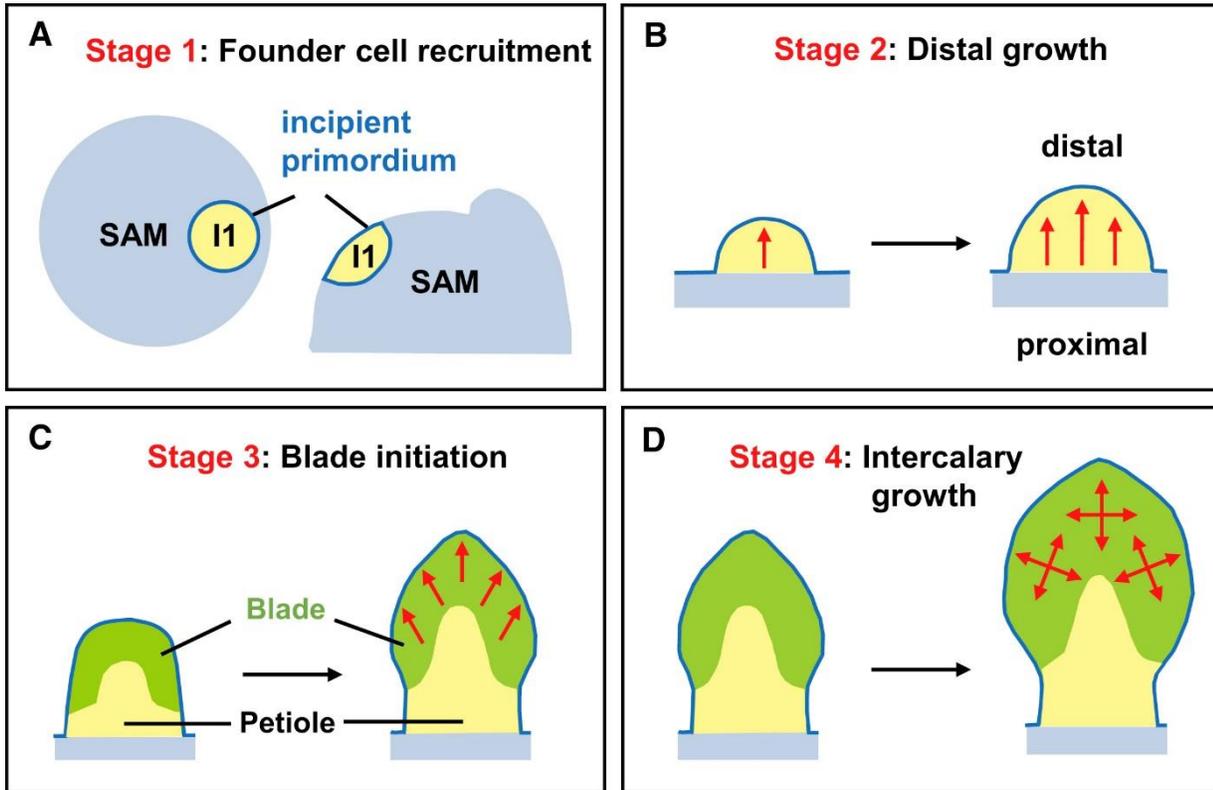


Figure 5: **Leaf morphogenesis**. Schematic representation of growth stages (A to D) encompassing leaf morphogenesis. Recruitment of founder cell from shoot apical meristem (SAM) (A) is followed by distal growth (B). At the margins of leaf primordium, the blade and petiole are specified (C). Marginal meristem activity stops, and subsequent events of cell proliferation and expansion sets the pace for both distal and lateral leaf growth (D). I1: the oldest incipient leaf primordium, top (left) and front (right) view (A). Figure reproduced after (Du et al., 2018).

Recently, based upon cellular localization studies and microtubular interactions, plant-specific IQ67 DOMAIN (IQD) proteins (Abel et al., 2005) are also suggested to be involved in the growth of plant organs possibly in $Ca^{2+}$ dependent manner during cytokinesis in *Arabidopsis* (Burstenbinder et al., 2017). Another crucial aspect of leaf organogenesis, as indicated above, is the cell to cell communication. Intercalary leaf growth envisages leaf blade formation, cell proliferation and differentiation (Nakata and Okada, 2013). Switching of cell state from proliferation to differentiation occurs via the formation of an imaginary and conceptual 'cyclic arrest front' that

distinguishes differentiating cells at distal end of leaf from those of meristematic or proliferating cells in proximal region of leaf thereby forming a dynamic basipetal gradient during leaf morphogenesis (Nath et al., 2003). Originally, it was speculated that cell proliferation 'arrest front' forms and remains at a fixed position from the leaf base and moves rapidly towards the base as leaf maturation takes place. However, this view has been challenged. In fact detailed analysis showed the 'cyclic arrest front' or the border between meristematic and non-meristematic domains forms and disappears somewhat rapidly without moving towards the base at a fixed position from leaf base (Kazama et al., 2010; Andriankaja et al., 2012). Recent studies shows, unlike originally proposed movable nature of "cyclic arrest front", that there exist some uncharacterized non-autonomous cell-cell communication modules which might link the extent of cell expansion to cell proliferation in leaf primordium (Kawade et al., 2010). Though, regulatory factors responsible for the regulation of 'arrest front' itself are still invasive, a complete molecular understanding of regulation of leaf organogenesis still unclear.

Apart from *mir396-GRF* and *mir319-TCP* regulatory modules, overexpression studies for *AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE* (*ARGOS*) has shown to affect the leaf size by its action upon the DNA-binding protein ANT (AINTEGUMENTA) and CYCD3;1. Ectopic expression of *ARGOS LIKE* (*ARL*) and *ORGAN SIZE RELATED1* (*ORS1*), which possess a common conserved domain, positively affects cell division and expansion in the leaf as demonstrated by alterations in the number and size of cells in leaf lamina (Hu et al., 2003; Feng et al., 2011). *AN3* and BR biosynthetic gene *ROTUNDIFOLIA* (*ROT3*) were some of the key factors discovered in late nineties to affect leaf shape in genetic exploration of leaf determinants (Tsuge et al., 1996). Lateral and longitudinal cell expansion has been shown to be regulated by *AN3* and *ROT3*, respectively. *ROT4,* that encodes a small peptide, which localizes in the plasma membrane without signal for secretion, and have 23 paralogs (*ROT-FOUR-LIKE* (*RTFL*)/*DEVIL* (*DVL*)), (Tsukaya, 2013a) is another candidate gene and it was found that its overexpression produces smaller shunted leaves and short stems (Narita et al., 2004; Ikeuchi et al., 2011).

In addition to the role of above described gene regulatory networks in leaf morphogenesis, phytohormones are also crucial for morphogenesis in plants and have versatile roles. Alterations in levels and/or associated downstream signaling of terpenoid-derived gibberellins (GAs) or BRs drastically affects leaf morphology. Transgenic plants carrying constitutive overexpression or mutations in genes encoding for biosynthetic enzymes for GAs or BRs pathways enhances leaf size while blockage of signaling or biosynthesis results in opposite phenotypic effects (Huang et al., 1998; Choe et al., 2001; Achard et al., 2009; Zhiponova et al., 2013). The downstream

signaling effects of GAs are initiated upon degradation of inhibitory DELLAs proteins that remains in bound form with GAs receptor, *GIBBERELLIN INSENSITIVE DWARF1* (*GID1*), in the absence of GAs (Harberd, 2003; Achard and Genschik, 2009; Daviere and Achard, 2013). Enhanced cell proliferation by GAs during intercalary growth is potentially achieved by via the inhibition of KIP-RELATED PROTEIN 2 (KRP2) and SIAMESE that are negative regulators of cell cycle (Achard et al., 2009). Various major plants hormones viz: auxin, GAs, CKs, ethylene, abscisic acid, jasmonic acid (JAs) and BRs show overlapping physiological functions due to shared gene networks. However, the actions of one plant hormone may not be necessarily countered by the loss or application of other plant hormone (McSteen and Zhao, 2008). Similarly, BRs effects are indispensable and have been shown to affect a wide range of cellular processes such as division, elongation and differentiation including the control of exit from mitosis (Zhiponova et al., 2013) during entire period of growth of various plant organs. BRs bind to its heterodimeric co-receptor complex, comprising BRASSINOSTEROID INSENSITIVE1 (BRI) and BRI1-ASSOCIATED RECEPTOR KINASE1 (BAK1) subunits, leading to dissociation of BRI1 KINASE INHIBITOR1 (BKI1) and simultaneous *trans*-phosphorylation in the kinase domains of BRI1 and BAK1. These events paves the way for the downstream signaling cascades via other factors (Vert et al., 2005; Zhu et al., 2013; Planas-Riverola et al., 2019). Other than the above mentioned role of auxin in establishment of leaf polarity during emergence of primordium, its function in leaf growth is subject to debate and indicated to have roles in cell division, enlargement, and differentiation (Perrot-Rechenmann, 2010). Perception of auxin in the nucleus triggers transcriptional responses by involving three core components, the ARFs, auxin co-receptors, F-BOX TRANSPORT INHIBITOR RESPONSE 1/AUXIN SIGNALING F-BOX PROTEIN (TIR1/AFB) and short lived transcriptional repressors, AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA). Auxin binding results in expedite interaction between TIR1/AFB and Aux/IAA proteins. Aux/IAA are degraded by proteasomal machinery which results in the release of ARFs and onset of associated gene expression responses (Lavy and Estelle, 2016; Luo et al., 2018). Recent reports highlighted another transcriptional module for auxin signaling by showing that auxin induces expression of the *miR847* that targets *IAA28* mRNA. Overexpression of *mir847* or *IAA28* knockout results in larger rosette leaves suggesting positive role of *mir847* in meristematic competence and in the determination of the duration of cell proliferation and lateral organ growth in *A. thaliana* (Wang and Guo, 2015). However, other contradictory report highlighted a negative role of polar auxin efflux transporter, serine/threonine kinase (*PINOID*), in cell proliferation and expansion (Saini et al., 2017).

Besides hormonal and gene regulatory modules, mechanical forces are also key determinants in control of the overall expansion of plant cells. Dynamics of cell wall mechanics and turgor pressure, a centrifugal force against the cell wall, determines direction and rate of cell expansion thereby final pattern formation in plant organs (Cosgrove, 2005; Dumais, 2007). Over the time due to increase in the elasticity, the cell wall in PZ becomes comparatively loose than those of CZ cells (Milani et al., 2011; Kierzkowski et al., 2012). *EXPANSINS* (*EXPs*) were discovered in 1992 as the nonenzymatic cell-wall-loosening proteins that reduces adhesion between adjacent and cross-linked wall polysaccharides, viz: cellulose, hemicelluloses and pectin, in a pH dependent manner that results in an increased cell wall extensibility (McQueen-Mason et al., 1992). In *A. thaliana, EXPANSINS* family consists of 36 members grouped into four distinct classes of α, β, Expansin-Like Family A and Expansin-Like Family B types (Lee et al., 2001). Studies shows that *EXPs* are emerging candidates that positively regulate organ development and plant cell enlargement by affecting the cell wall properties (Marowa et al., 2016). Other than cell proliferation and expansion, change in ploidy state of cell has also been shown to affect cell size. Replication of genome without mitosis results in change of ploidy, a process that is termed endoreplication or endoploidization. Endoreplication has also been found to positively affect organ size by increasing size of cells. In contrast, change in cell number rather than the size of cell has been advocated to predominantly affect final dimensions of leaf (Gázquez and Beemster, 2017). Genetic studies employing genes from different functional classes showed that increased leaf size due to the overexpression of *ARABIDOPSIS THALIANA V-PPASE 3* (*AVP1)*, *GRF5*, *JAGGED AND WAVY* (*JAW)*, *BRI1*, and *ARABIDOPSIS THALIANA GIBBERELLIN 20-OXIDASE 1* (*GA20OX1)* occurs mainly due to increase in cell number (Gonzalez et al., 2010). Additionally it was reported that mutants showing change in number of cells tends to compensate the phenotype by alterations in size of cells. For instance overexpression of *AN3* increases leaf area by an increase in cell numbers while the *an3* mutant exhibits smaller leaves with reduced number of enlarged cells that are twice in size compared to WT, pinpointing to the central role of cell proliferation and expansion in final leaf size (Horiguchi et al., 2005; Hisanaga et al., 2015). This phenomenon is called 'compensation' and play vital role in organ size determination (Tsukaya, 2002a).

Additionally, environmental factors are also crucial for final leaf shape and size. Plant physiology is dependent over the photosynthesis, so is the adaptation of leaf shape to the light direction and intensity. Leaves undergoing growth in weak light condition tend to elongate petioles as an adaptive response to capture light. This phenomenon is terms as 'shade avoidance' and

suggested to be modulated by interplay of helix-loop-helix transcription factors called *PHYTOCHROME INTERACTING FACTORS* (*PIFs*) and the photoreceptor genes *PHYTOCHROME A* to *E* (*PHYA* to *PHYE*) where *PHYA* and *PHYB* are considered as the most important regulator of light response (Leivar and Quail, 2011; Casal, 2012). Certainly, the effect of light intensity over the thickness of mesophyll was shown long time ago as a physiological productivity (Björkman, 1981). Recent advances highlighted that PHOTOTROPIN 2 (PHOT2), a photoreceptor, increases the length of mesophyll cells in palisade layers in the direction of leaf thickening in response to the light to maximize photosynthetic efficiency (Kozuka et al., 2011; Gotoh et al., 2018).

Though several genes can directly or indirectly impact mesophyll size, *mir396-GRF/GIF* component is often cited as a more direct regulator of cell proliferation for mesophyll cells among others (Ren et al., 2019). In fact, the finding that *AN3* is specifically expressed in mesophyll cells and further moves as a protein within layers of other cell types highlights importance of mesophyll layer as a key signaling source for synchronized cell proliferation in a leaf primordium (Kawade et al., 2013). Other potential genes that could be directly involved in regulation of mesophyll cell morphology are *AINTEGUMENTA* (*ANT*) (Mizukami and Fischer, 2000), *Arabidopsis SKP1-LIKE1* (*ASK1*) (Zhao et al., 1999), *NGATHA* (*NGA*) (Lee et al., 2015), *G-PATCH DOMAIN PROTEIN1* (*gdp1*) (Kojima et al., 2017), *KLU* (Anastasiou et al., 2007), *OLIGOCELLULA1, 4, 6* (*oli1, 4, 6*) (Fujikura et al., 2009), *POINTED FIRST LEAD 2* (*PFL2*) (Ito et al., 2000), *ROTUNDIFOLIA4* (*ROT4*) (Narita et al., 2004), *STRUWWELPETER* (SWP) (Autran et al., 2002), and *TCP4* (Schommer et al., 2014). In summary, the final organ size ultimately depends on the ability of plants to integrate external and internal cues to balance and optimize organ growth via temporally and spatially regulated gene networks. Nevertheless, a complete understanding of leaf development is yet to be achieved. The identification of the action mechanisms of gene(s) that influence leaf growth can surely contribute to the advancement of our overall knowledge of plant biology and potential agricultural applications.

# 3  Thesis objectives

In order to get insights into the potential function and mechanism of action for *lncNATs-UGT73C6* previous studies were performed in the laboratory. As mentioned earlier (section 2.2.3) *lncNATs-UGT73C6* share high level of sequence complementarity with other members of *UGT73C* sub family (Table 1). It was hypothesized that *lncNATs-UGT73C6* can form dsRNAs not only with fully complementary mRNAs of *UGT73C6* but also with the ones from *UGT73C5* and other *UGT73C* subfamily members if co-expressed. In agreement with this hypothesis, co-overexpression of *lncNATs-UGT73C6* with *GFP*-tagged sequences of *UGT73C6* subfamily members in the model plant *N. benthamiana* leading to reduction of GFP signal (Figure 6)(de-Vries, 2014; Fritz, 2015), a result that was confirmed by western blot (de-Vries, 2014; Fritz, 2015). Additionally, co-expression resulted in the production of small RNAs, an observation that prompted us to hypothesize that *lncNATs-UGT73C6* can downregulate not only *UGT73C6* but also can form locus-specific regulatory loop with other family members.
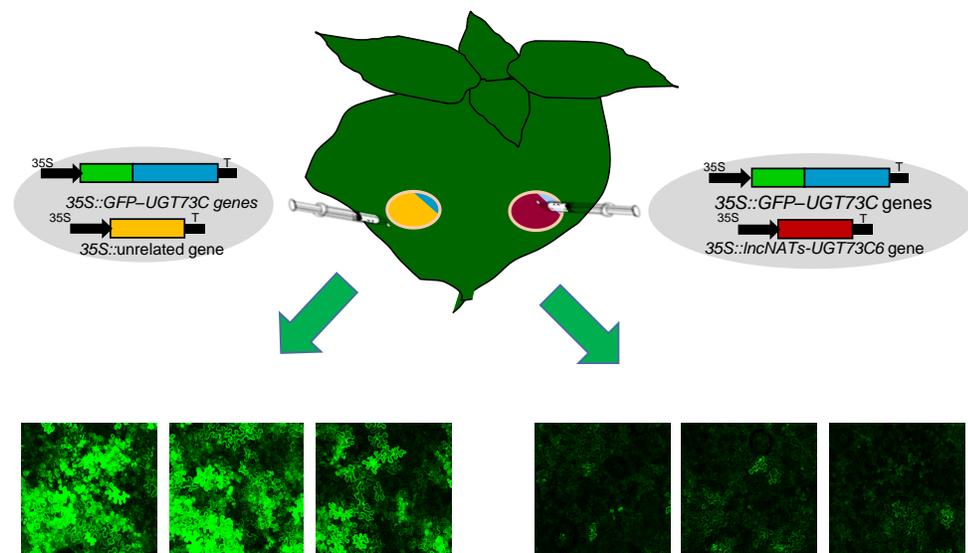


Figure 6: **Transient co-expression of *lncNATs-UGT73C6* with *UGT73C* family members results in their downregulation**. Representative schematic showing transient co-expression of *lncNATs-UGT73C6* and *UGT73C* family members in *Nicotiana benthamiana*. Left and right side from *N. benthamiana* leaves were co-infiltrated with *35S::GFP-UGT73C* genes with 35S::unrelated genomic sequence and *35S::lncNATs-UGT73C6* independently. 48-72 hours post infiltration (hpi) leaf discs from infiltrated sites (squires in sketched leaf) from both left and right side were checked under confocal microscope for expression of GFP. Leaf discs carrying co-expression of both *lncNATs-UGT73C6* and *UGT73C* genes shows decreased GFP signal intensity (bottom right) compared to control (bottom left). Results were confirmed by western blot. Figure is modified after (de Vries, 2014).

Additional sets of preliminary experiments were also performed to analyze if *pORFs* included in the *lncNAT2* sequence (Figure 7) undergo translation to assess whether *lncNATs-UGT73C6* are bona fide lncRNAs. In order to test the potential peptide coding capacity of *lncNAT2*, each *pORF* was fused to *GFP* and the constructs were transiently expressed in *N. benthamiana*. Interestingly, all *pORFs* with the exception of *pORF4* were translated, producing GFP-tagged peptides. Moreover, product of *pORF3,* which has stretches of basic amino acids, showed specific localization to nucleolar bodies suggesting that this peptide could play role in the nucleus (Figure 7) (de-Vries, 2016).
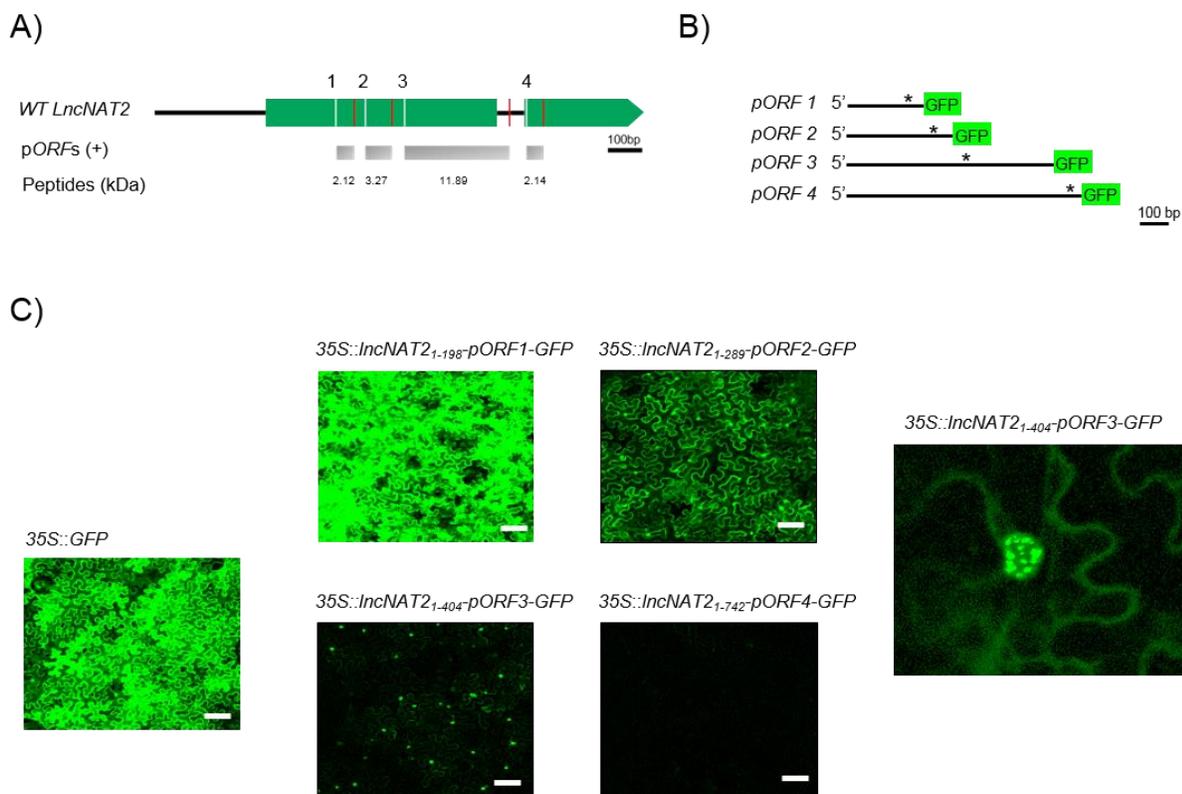


Figure 7: **The first three *pORFs* present in *lncNAT2* are translated when overexpressed in *N. benthamiana*.** A) Genomic organization of *lncNAT2 pORFs*. B) Respective *pORFs* sequence is fused to GFP after removing stop codon (indicated by asterisk). C) *pORFs* are overexpressed under control of 35S promoter. Leaves of *N. benthamiana* were infiltrated with *Agrobacterium tumefaciens* carrying respective plasmid for control GFP and *pORFs*. GFP signal in leaf discs from infiltrated area was visualized by confocal microscopy post 48 hpi. Magnified image (far right) shows localization of the product of *pORF3* in the nucleus. (Figure is adapted after Vries, 2016).

*LncNAT1* and *lncNAT2* reporter β-glucuronidase (GUS) lines were previously generated by cloning 2500 and 2000 bp upstream of the annotated transcription start site (Source : TAIR 10) of *lncNAT1* and *lncNAT2* respectively (de-Vries, 2014). Analysis of reporter activity for *lncNAT1_{prom}::GUS* and *lncNAT2_{prom}::GUS* highlighted that both *lncNAT1* and *lncNAT2* promoters are active in *A. thaliana.* Furthermore, *lncNAT1* reporter GUS activity displayed overlapping spatiotemporal expression with *UGT73C5_{prom}::GUS* in roots (Fritz, 2015) while *lncNAT2_{prom}::GUS* showed partially overlapping expression pattern with the reported *UGT73C6_{prom}::GUS* activity in seedlings (Husar et al., 2011). Consequently, reporter GUS activity analysis could only partially support the preliminary results regarding the potential of *lncNATs-UGT73C6* to regulate expression of *UGT73C* subfamily members as mentioned in the previous paragraph. Yet, the detection of the expression and _{prom}::GUS activity of *lncNATs-UGT73C6* in different tissues and developmental stages in *A. thaliana* paved the foundation to continue and adopt a multidirectional comprehensive approach for further investigations.

On one hand, transient expression of *lncNATs-UGT73C6* with *UGT73C* family members results in their downregulation potentially by small RNAs while at the same time transient infiltration studies highlighted that *pORFs* can produce peptides to achieve specific functions indicating that *lncNATs-UGT73C6* might not act as bona fide lncRNAs. Based upon these observations in the non-host plant *N. benthamiana,* we sought to investigate the function of *lncNATs-UGT73C6* in real host plant *A. thaliana*. Although a few NAT-lncRNAs are characterized and shown to regulate fundamental developmental processes in *A. thaliana* (section 2.1.3), studies exemplifying biological roles of NAT-lncRNAs particularly from multigene families in plants are still in infancy. Therefore, to gain insight into the roles of NAT-lncRNAs from multigene families, in their host plant, the study was aimed at the *de novo* functional and molecular characterization of annotated *lncNATs-UGT73C6*. The following objectives have been attempted to address the role of *lncNATs-UGT73C6* in *A. thaliana*.

1. **To characterize *lncNATs-UGT73C6* at the molecular level and to evaluate the phenotypic effects of alteration of their expression**
2. **To identify whether *lncNATs-UGT73C6* form a local gene expression regulatory loop with the sense gene *UGT73C6* or other closely related *UGT73C* family members.**
3. **To investigate the function, target molecules and potential mechanism of *lncNATs-UGT73C6* action.**

# 4 Methods

## 4.1 Amplification, cloning and transformation

Desired DNA fragments were amplified using standard PCR program with an initial denaturation step at 95-98 °C for 30 seconds followed by 35 cycles consisting of denaturation at 95-98°C for 15-30 seconds; annealing based on the melting temperature of the primers for 10-20 seconds and extension at 72°C for 15-30 seconds per kb depending upon the type of polymerase used. A final extension was performed at 72°C for 5-7 minutes. PCR reactions were performed by using either Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) or DreamTaq DNA Polymerase (Thermo Fisher Scientific) as per manufacturer protocol. Forward primers used for directional cloning into pENTR™/D-TOPO vector (2580 bp, Thermo Fisher Scientific) contained an additional CACC sequence at the 5'; end.

Purified DNA amplicons were further cloned into pENTR™/D-TOPO vector (2580 bp, Thermo Fisher Scientific) following manufacturer's instructions and transformed into homemade competent DH5α or One Shot® TOP10 (Thermo Fisher Scientific) *E. coli* cells. Competent bacteria were heat shock transformed (1 min incubation at 42 °C), recovered using SOC media without antibiotics, incubated 1 h at 37 °C with agitation, plated on LB agar plates in presence of 50 µg/ml (Kanamycin) and growth overnight at 37 °C. Each insert cloned into the pENTR™/D-TOPO vector was verified by DNA sequencing of plasmids extracted from positive clones. Insert of selected plasmid were cloned by gateway cloning into binary destination vectors (Karimi et al., 2005) using Gateway® LR Clonase™ II enzyme mix (Invitrogen®). DH5α cells were transformed by electroporation, recovered in media without antibiotics, incubated during one hour with agitation at 37 °C, and selected by overnight growth at 37 °C in LB media plates containing 75 µg/ml Spectinomycin. Isolated single colonies containing recombinant plasmids were further grown in liquid LB medium that was supplemented with suitable antibiotics for the purpose of plasmid isolation.

GV3101 strain of *Agrobacterium tumefaciens* was transformed by electroporation with the selected pDEST plasmids. Transformed *A. tumefaciens* were recovered in media without antibiotics, incubated at 28 °C during two hours with agitation, and selected on LB containing plates by incubation at 28-30°C for 48 hours in presence of Rifampicin, Gentamycin and Spectinomycin at final concentrations of 25 µg/ml, 10 µg/ml and 75 µg/ml, respectively. Identity of plasmids was verified by colony PCR. Subsequently *A. tumefaciens* carrying different constructs

were used for floral dipping transformation of *A. thaliana* Columbia-0 (*Col-0*) ecotype for stable transformation.

## 4.2 Generation of transgenic lines

### 4.2.1 Reporter GUS and translation fusion constructs

*UGT73C6* and *UGT73C5* reporter GUS lines were generated using the similar promoter region as described earlier (Poppenberger et al., 2005; Husar et al., 2011). Briefly, in order to generate stable $_{prom}$::GUS lines for *UGT73C6*, 1687 bp genomic sequence upstream of ATG start site was PCR amplified from genomic *A. thaliana* DNA by Phusion® polymerase (Thermo Scientific). Subsequently the purified PCR product was directionally cloned into pENTR™/D-TOPO vector (section 4.1). Further sequencing confirmed that cloned product was in frame with GUS in the Gateway® destination vector pBGWFS7 (Karimi et al., 2005). Similar cloning procedure was adopted for the generation of *UGT73C5* promoter GUS lines that contained 760 bp sequence upstream of ATG start site fused with reporter GUS.

The translational fusion lines for *UGT73C6* and *UGT73C5* were generated by cloning complete open reading frames (ORFs), 5' untranslated regions (UTR) and above mentioned upstream promoter sequences of *UGT73C6* and *UGT73C5* into the binary destination vector pBGWFS7 that contains the sequences encoding the green fluorescent protein (GFP) and GUS. Sequencing was performed after Gateway® cloning to confirm that the ORFs are in frame with GFP-GUS.

Subsequently, selection of lines until $T_3$ homozygous stage from $T_1$ generation was carried out using resistance to herbicide phosphinothricin (BASTA) as selection marker. $T_2$ to $T_3$ screening of transgenic lines was carried out following 3:1 Mendelian segregation, indicative of single insertion. Homozygosity of $T_3$ lines were confirmed by complete resistance (survival of 100% of the seedlings) on ½ MS agar plates. Following experiments were carried out using $T_3$ seeds derived from homozygous plants.

### 4.2.2 Overexpression and artificial microRNA (amiRNA) downregulation constructs for *lncNATs-UGT73C6*

35S overexpression constructs for *lncNAT1* and *2* were generated by cloning genomic DNA sequence of *lncNATs-UGT73C6* that corresponds to annotated transcripts of *lncNAT1* and *lncNAT2*. Genomic fragments for *lncNATs-UGT73C6* were PCR amplified and cloned into

pENTR™/D-TOPO, and subsequently sub-cloned in the pB7WG2 vector (Karimi et al., 2005) as per Gateway® cloning procedure described in section 4.1. Subsequent to cloning in *E. coli* and *A. tumefaciens transformation*, WT *Col-0* plants were transformed by floral dipping and transgenic lines were selected as described in 4.2.1.

Artificial microRNA (amiRNA) lines were generated to downregulate *lncNATs-UGT73C6* in *trans*-specific manner by employing a web based P-SAMS amiRNA designer tool and following the cloning procedure reported earlier (Carbonell et al., 2014). Briefly, for direct cloning of amiRNA insert into AtMIR390-B/c vector, sense and antisense oligonucleotides (Table 7) were annealed using oligo annealing buffer (0 mM Tris-HCl (pH 7.5), 500 mM NaCl, 60 mM MgCl$_2$, 10 mM DTT). Subsequently, digestion-ligation reaction with AtMIR390-B/c vector (kindly provided by Dr. Carbonell) was carried by employing diluted annealed oligonucleotides, T4 DNA ligase, T4 DNA ligase buffer and *BsaI* following heat shock transformation of Ccdb (toxin) susceptible *E. coli* strain One Shot® TOP10 (Thermo Fisher Scientific®) as described in section 4.1. The transformants were analyzed by PCR and sequencing. Three independent amiRNAs constructs were generated to target both *lncNATs-UGT73C6* at different locations (Figure 11 A).

Subsequent to cloning procedure and transformation of *A. tumefaciens*, WT *Col-0* plants were transformed by floral dipping with the amiRNAs constructs. Subsequently, selection of lines until T$_3$ homozygous stage from T$_1$ generation was carried out using Hygromycin B as selection marker on plates as described in 4.2.1. Expression and presence of mature amiRNAs targeting *lncNATs-UGT73C6* RNA were confirmed by Northern blot (section 4.11) using $^{32}$P $\gamma$-ATP labelled oligo probes (Table 7).

## 4.2.3 Site directed mutagenesis of *lncNAT2* open reading frames

Primers were designed for the site directed mutagenesis of start codons of *pORFs* present in *lncNAT2* sequence (Table 7). ATG start site codons of *pORFs* were mutagenized into CCC by using *Accu*Prime™ *Pfx* DNA Polymerase (Life Technologies®) following manufacturer's instructions. pENTR™/D-TOPO plasmid containing the *lncNAT2-UGT73C6* construct mentioned in section 4.2.2 was used as template for the PCR reaction. Non-mutated parental plasmid was eliminated by *Dpn* I digestion and the reaction mixture was used for transformation of competent cells. Subsequently positive clones were verified by colony PCR, used for gateway cloning into pB7WG2 (Karimi et al., 2005), multiplied in *E. coli* and transformed into *A. tumefaciens*. Agrobacteria containing the constructs were used to transform *A. thaliana* by floral dip.

Independent $T_3$ homozygous lines for *Mut_lncNAT2* were selected following Mendelian segregation and selection process using BASTA herbicide as mentioned in section 4.2.1.

## 4.3  CRISPR/Cas9 editing of *lncNATs-UGT73C6*

Clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) technology was adopted to specifically knockout *lncNATs-UGT73C6* using the double guide RNA (gRNA) approach[1]. The vectors were obtained from Dr. Mily Ron (University California, Davis) and the site specific gRNAs (Table 7) were generated using CHOPCHOP webserver (http://chopchop.cbu.uib.no/). PCR reactions were performed using appropriate primers combinations, Phusion DNA polymerase (Thermo Fisher Scientific®) and scaffold plasmid as template. Digestion-ligation reactions were performed by employing T4 DNA ligase and *BstI,* the purified PCR product and pENTR vector (L1L2-AtU6gRNA)*.* One Shot® TOP10 *E. Coli* cells were heat shock transformed with reaction mixture and positive colonies were identified. Subsequent cloning reaction was performed using Clonase II (Thermo Fisher Scientific®) according to manufacturer instructions into pMR333 (CRISPR/Cas9), a destination vector that expresses green fluorescent protein (GFP*)* under the control of seed coat promoter (*OLE1:: GFP*) and CRISPR endonuclease caspase9 (Cas9) from a constitutively active ribosomal promoter (*RPS5A::Cas9).* Plasmid was transferred in *A. tumefaciens* by electroporation followed by transformation of *Col-0* as described earlier (section 4.1). $T_1$ selection of positively transformed seeds was carried by visual inspection of seeds that expressed GFP in seed coats using Leica MZ FLIII fluorescence stereomicroscope (Leica Fluo Combi™). Subsequent screening of lines at $T_2$ stage was carried out by choosing non-GFP expressing seeds confirming the absence of pMR333 vector in selected seeds. Absence of *Cas9* in $T_2$ plants was further validated by PCR using Dream Taq polymerase (Thermo Scientific) followed by sequence verification of allelic state in individual plants that carry intended deletions for *lncNATs-UGT73C6*. All experiments were performed using seeds derived from homozygous plants. Primers used for assembly product are listed in Table 7.

## 4.4  Plant material and growth conditions

Surface sterilized (10% v/v bleach and 0.05% Triton X-100) seeds of WT *Col-0* and transgenic lines were stratified for 2 days at 4 º C followed by germination and growth over ½ MS plates under 16 hours light and 8 hours dark cycle. Unless mentioned specifically, for majority of phenotypic experiments 10 days old seedlings were transferred to soil in 8cm round pots

---

[1] This work was performed together as a team with Mr. Ammar Jaber

containing steam-sterilized CL clay supplemented with phyllosilicate mineral and vermiculite (EINHEITS ERDE®). Control and transgenic plants in growth cabinets (Clf Plantclimatics GmbH) were randomized within the trays that were rotated periodically and were irrigated with fixed water volume under tightly controlled conditions of light (95-110 µmol/m$^2$), temperature (18ºC night and 21ºC day) and humidity (65-70%). Every individual phenotypic experiment included between 15 and 30 plants per transgenic line. Each tray contained minimum 2 plants from chosen control and transgenic lines. Plants grown under greenhouse conditions were controlled only for irrigation unless specified.

## 4.5  Analysis of phenotypic parameters

### 4.5.1 Root length and biomass measurements

Appropriately spaced seedlings were grown over ½ MS 125 x 125 mm square vertical plates for 14 days and were subjected to digital image acquisition by using a desktop scanner (HP ScanJet Co.) at 300 dots per inch (dpi). The output images were stored as TIFF (tagged image file format) files. Originally scanned TIFF images were then transformed into grayscale 8 bit images and further analysis performed by freely available National Institute of Health (NIH) Image J software (Collins, 2007; Schneider et al., 2012) that was equipped with updated neuron J plugging (Meijering et al., 2004). The 8 bit images were uploaded onto the Image J program and after fixing scale value (in pixels) the primary roots were traced using a segmented line tool followed by measure command for calculation of root length.

For biomass measurement intact 14 days old individual seedlings, grown in condition as mentioned for root measurement, were weighed using a high precision weighing balance (Sartorius®). Individual weights were recorded for 40 seedlings per genotype.

### 4.5.2 Quantification of phenotypic traits

For the estimation of complete rosette and individual leaf area, high quality plant pictures were recorded with a manually fixed focal length using a Sony® DSLR camera at different time points. At 25 days care was taken to avoid superimposition of leaves while at 35 days the stems and cauline leaves were clipped off before making pictures. The digital images were than loaded onto the Easy Leaf area® program (Easlon and Bloom, 2014) and green leaf area was quantified using a pre-specified red squire as scale according to the program manual (Supplementary Figure

3). Other leaf parameters such as timing for emergence of inflorescence, opening of first flower, stem height, flowering time, leaf number and leaf dimensions were calculated manually.

In order to quantify total seed yield, plants were grown in greenhouse or growth cabinets using Arasystem® 360 Kit consisting Aratube and Arabase. The system allowed to prevent any seed loss without hampering ventilation, access to light source and thus normal growth of individual plants. Seeds were carefully collected after the completion of senescence stage and weighed using high efficiency automatic weighing balance (Sartorius ®).

### 4.5.3 Analysis of leaf cell size and area

$6^{th}$ leaf in each individual plant was marked soon after emergence. In addition to rosette area quantification at 25 days, the plants at 35 days were quantified for rosette area estimation after dissecting the stems and cauline leaves. Statistical analysis was performed from the overall rosette area values to mark the individual plants that were closer to median values in individual groups. From these two plants, $6^{th}$ leaves were subjected for histological preparation according to earlier reports (Nelissen et al., 2013; Wang and Guo, 2015). The transparent intact leaves were divided into four segments of equal dimensions (length and width) from bottom to tip after dissecting petiole. Mid-rib and adjacent region from both left and right fronts of each leaf discs was cautiously removed followed by brief incubation and mounting of leaf discs in Hoyer medium and lactic acid respectively (Nelissen et al., 2013). Using differential interference scanning (DIC) mode in ApoTome 2 (ZEISS) microscope, 3-5 pictures from each respective leaf specimen were recorded at 10 x magnification for adaxial epidermis and palisade mesophyll cells. A minimum of 8 to 12 total pictures from both leaves were selected from bottom and tip region and analysis of cell size and area was performed using Image J program (Collins, 2007; Schneider et al., 2012) and procedure as described before (Nelissen et al., 2013; Wang and Guo, 2015).

## 4.6  Histochemical GUS assay

Transgenic lines carrying GUS reporters were subjected to GUS staining. Seedlings or leaf material were incubated at 37º C in GUS staining solution that contains the substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc). Consequent upon staining, the plant material was serially dehydrated to remove chlorophyll using a series of 20 %, 35 %, 50 % and 70% ethanol solutions to visualize blue color staining that indicates GUS activity in plant organs. The pictures were recorded using NIKON SMZ1270® stereomicroscope magnifier.

## 4.7  Gene expression analysis by qRT-PCR

Total RNA was extracted from plant material using PEQlab RNA extraction Kit following manufacturer's instructions. Subsequent, expression analyses for selected genes were performed following reverse transcription by Superscript II® (Invitrogen) to synthesize complementary DNAs and quantitative real time PCR (qPCR) using Quantstudio 5 (Applied Biosystems™) thermal cycler.  Strand specific cDNAs were prepared using gene specific primers for *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* including the reverse primer for the reference gene in each cDNAs synthesis reaction. Other genes were analyzed using oligo d(T) primer for cDNA synthesis. qPCR reaction of 10 µl volume consisted 5 µl SYBR premix (Applied Biosystems), 1 µl undiluted cDNA, 0.1 µl forward and reverse primers at the concentration of 10 pmol/µl and 3.8 µl sterilized water. qPCR reaction cycle included initial denaturation at 94º for 30 s followed by standard 40 cycles of 94º C for 10 s, 60º C for 30 s and final melting curve determination that rise from 60º C to 94º C with gradual increments of 0.5 ºC each 15 s. The qPCR reaction was performed using 0.1 ml MicroAmp™ Fast Optical 96-Well reaction plate (Thermo Scientific®) using fast mode in qPCR system. All qPCR reactions included a reference gene, negative controls for cDNA and water as non-template control for qPCR reaction. Average Ct values were obtained from a minimum of 2-3 technical replicates for three individual biological replicates and were normalized by appropriate reference gene (*PP2A* in most of the analyses except *GAPC2A* in BR response assay). Final calculations were carried out by $2^{-\Delta\Delta ct}$ (adjusted to primer efficiency) method where fold changes in expression levels of genes are relative to reference gene *PP2A* or *GAPC2A*. Data presented are representative of 2-3 experiments. Each individual experiment included 3 biological replicates.

## 4.8  RNA stability assay

RNA stability assays were performed[2] according to previous report (Fedak et al., 2016). Briefly, 11 days old *Col-0* seedlings grown in ½ MS media containing plates were transferred to incubation buffer (1 mM PIPES at pH 6.25, 1 mM trisodium citrate, 1 mM KCl and 15 mM sucrose) for 30 minutes followed by addition of 150 mg/l cordycepin® (3'-deoxyadenosine, Sigma Aldrich) and vacuum-infiltration twice for 5 min. Seedlings were collected and total RNA was isolated  as described before at 0, 15, 30, 45, 60, 75, 90, 105 and 120 mins post cordycepin treatment. cDNA was synthesized using oligo d(T) and Superscript II® (Invitrogen) and qPCR measurements were

---

[2] These experiments were performed by Ms. Susane Engelmann

performed using gene specific primers (Table 7). *EXPANSIN-LIKE1* (*Expansin L1*), a short-lived transcript, and *EUKARYOTIC TRANSLATION INITIATION FACTOR 4A1* (*EIF4A1A*), a long-lived mRNA, were used as positive controls for cordycepin assay (Golisz et al., 2013). Normalization of Ct-values was performed by selecting Ct-value at time point 0 as 100 percent. A degradation curve was obtained from Ct values and slope of the curve was calculated using formulae (1) and (2) for the determination of half-life ($t_{1/2}$) of *lncNATs-UGT73C6* transcripts and above mentioned control genes.

$$Ct(n) = (ln(Ct/Ct(0))*(-10) \qquad (1)$$

$$t_{1/2} = (ln2)/|slope| \qquad (2)$$

## 4.9  Brassinosteroid treatment assay

Similar to BR assays reported in previous publications (Husar et al., 2011), 5 days old *Col-0* seedlings were grown vertically in squire plates containing ½ MS and 1.5 % agar. Afterwards seedlings were transferred to liquid ½ MS media and allowed to acclimatize for 2 days under continuous light conditions (~130 µmol·m$^{-2}$·s$^{-1}$). 1 µM 24-epibrassinolide (24-Epi BL, Sigma Aldrich®) dissolved in dimethyl sulfoxide (DMSO), or the same volume of DMSO was applied in treated and untreated sample sets, respectively (Supplementary Figure 10 A). Total RNA was extracted from seedlings collected at different time points and reverse transcribed into cDNA using oligo d(T) and Superscript II (Invitrogen®). The qPCR analysis was performed for *lncNATs-UGT73C6*, *UGT73C6*, *UGT73C5* and assay control genes viz: *DWF4*, *BAS1* and *BR6ox2*. Primers used this study are listed in Table 7.

## 4.10 DON treatment assay

Similar to Epi-BL treatment assay, 5 days old seedlings were transferred to liquid media for acclimatization for 2 days under long day conditions. Subsequently seedlings were incubated for 4 hours with either 16.9 µM fungal toxin DON dissolved in ethanol or same volume of ethanol according to earlier study (Poppenberger et al., 2003) (Supplementary Figure 11 A). After treatment, the seedlings were subjected to total RNA extraction. cDNA synthesis was performed using strand specific primers. Subsequently qPCR expression analysis was carried out for *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* using primer listed in Table 7. Additionally, a set of seedlings (DON treated and untreated) grown in parallel were used for GUS staining according to the procedure mentioned in section 4.6.

## 4.11 Northern Blot

Total RNA from fresh leaf tissue was extracted using phenol-chloroform method following a modified protocol (Suzuki et al., 2004). Small RNAs were separated by electrophoresis on 17% polyacrylamide, 7 M urea and 0.5 X TBE gel while 1% agarose, 1 X TBE and 1 M urea gel was used for the separation of high molecular weight RNAs. The RNAs were transferred onto the nylon membranes (Roche ®) by using a semi-dry electrotransfer unit (BioRad®) and cross linked twice with UV. For the detection of small RNAs, labelling of oligonucleotides probes was performed using T4 Polynucleotide kinase reaction in the presence of $^{32}$P $\gamma$-ATP followed by purification of oligo probes with miniQuick oligo columns (Roche®). Detection of high molecular weight RNAs was carried out with radiolabeled RNA probes. For the preparation of $^{32}$P α-UTP labeled riboprobes, appropriate plasmids were linearized by suitable restriction enzyme digestion following extraction and precipitation by phenol/chloroform/isoamyl alcohol and ethanol respectively. Linearized plasmids were further used as template for the *in vitro* synthesis of radiolabeled riboprobes using T7 RNA polymerase (Roche®) following manufacturer instructions. Additionally, integrity and concentration of the riboprobes was further checked by performing denaturing polyacrylamide gel electrophoresis. Blotted membranes were pre-hybridized using Perfect Hyb™ Plus buffer (Sigma®) at 68º C followed by 16 hour hybridization with appropriate probes at 38-42 ºC (small RNAs) or at 68 °C (high molecular weight RNAs). The hybridization solution was carefully removed and membraned were subjected for a series of low and high stringency washing steps in SSC buffer supplemented with 0.1% SDS at suitable temperature. Membranes were exposed to X-ray films and the signal was visualized using Typhoon FLA 7000 phosphoimager. Oligo probes used for the specific detection of amiRNAs are listed in Table 7.

# 5 Results

## 5.1 *LncNATs-UGT73C6* are transcribed from the complementary strand of *UGT73C6* and their expression is developmentally regulated

Strand-specific reverse transcription followed by PCR (RT-PCR) confirmed that annotated *lncNAT1* and *lncNAT2* are expressed from the complementary strand of *UGT73C6* in *A. thaliana* (Figure 8 B). In order to examine whether detected *lncNATs-UGT73C6* transcripts originate from distinct promoter elements downstream of *UGT73C6*, β-*Glucuronidase* (*GUS*) reporter construct were generated in *Col-0* background for *lncNAT1* and *2* (de-Vries, 2014). The purported promoter region in *lncNAT1$_{prom}$*::*GUS* and *lncNAT2$_{prom}$*::*GUS* constructs included 2058 and 2585 bp sequence upstream of annotated transcription start site for the respective *lncNAT*. Reporter activity for *lncNAT1* and *lncNAT2* promoter showed a characteristic spatiotemporal expression pattern and separation of expression domain in roots and shoots respectively (Figure 8 A). Expression of *lncNAT2$_{prom}$*::*GUS* was observed at the bottom-middle region of leaf adjacent to petiole in nascent young leaves, a region that possess actively dividing and proliferating cells (Du et al., 2018). Time course examination of *lncNAT2$_{prom}$*::*GUS* activity shows a gradual decrease over time in staining intensity with the increasing age of leaf whereas *lncNAT1$_{prom}$*::*GUS* shows continuous basal expression in roots and remains absent in other tissues during different developmental stages analyzed in this study (Figure 8 A and Supplementary Figure 1)*.*

To examine if expression of *UGT73C6* and *UGT73C5* overlaps with *lncNATs-UGT73C6* in *A. thaliana*, we analyzed reporter activity using $_{prom}$*::GUS* lines. Previous work (Fritz, 2015) showed unexpected ubiquitous GUS activity of a *UGT73C6$_{prom}$::GUS* reporter line (provided by Prof. B. Poppenberger), which differed from the reported expression pattern of *UGT73C6* (Husar et al., 2011). Reporter GUS activity analysis of above mentioned *UGT73C6$_{prom}$::GUS* line altogether with our reporter *GUS* lines of *lncNATs-UGT73C6* showed that their spatio-temporal expression pattern overlaps throughout the plant organs including the leaves (Fritz, 2015). On one hand, this observation satisfied the condition of sense and antisense co-expression thereby potential nat-siRNA-mediated regulation of *UGT73C6* by *lncNATs-UGT73C6* in agreement with the results from transient expression analysis in *N. benthamiana* (Figure 6). However, it was inconsistent with the previously published results about the *UGT73C6$_{prom}$::GUS* expression pattern (Husar et al., 2011). To solve these discrepancies, we generated *UGT73C6$_{prom}$::GUS* constructs harboring 1687 bp sequence upstream of ATG start site in the construct (Husar et al., 2011). Additionally, we generated *UG773C5$_{prom}$::GUS* carrying a 760 bp sequence upstream of translation start site of

*UGT73C5* (Poppenberger et al., 2005). Analysis of the *UGT73C6$_{prom}$*::*GUS* lines indicated that, at 7 days after sowing (DAS), the promoter activity tends to confine to spot-like patches in the cotyledons and to roots, which shows a homogenous staining pattern (Figure 8 A). In comparison, *UGT73C5$_{prom}$::GUS* showed a stronger GUS activity than those of the reporter lines of *UGT73C6, lncNAT1* and *2* (Figure 8 A). Analysis of $_{prom}$*::GUS* lines at 7 DAS, showed that both *UGT73C5* and *UGT73C6* promoters do not show overlapping expression with *lncNAT2$_{prom}$*::*GUS* in the first true leaves (Figure 8 A). *UGT73C6$_{prom}$::GUS* remains poorly active after 5 DAS and shows only sporadic expression at the leaf serrations or no expression in the vegetative tissues throughout the development (Supplementary Figure 1). Interestingly, as the plant develops, *UGT73C5$_{prom}$::GUS* activity increases in leaf areas previously occupied by *lncNAT2$_{prom}$*::*GUS* activity (Figure 8 A and Supplementary Figure 1). Thus during the growth of *A. thaliana lncNAT2$_{prom}$*::*GUS* and *UGT73C5$_{prom}$::GUS* exhibit an inversely correlated expression pattern in leaves. This peculiar follow-up *GUS* expression pattern of *UGT73C5* after *lncNAT2* is present in each of new emerging leaf. Analysis of $_{prom}$*::GUS* lines at 14 DAS also shows that *lncNAT2* and *UGT73C5* expression is spatiotemporally separated during leaf morphogenesis. Thus, over the time *UGT73C5$_{prom}$*::*GUS* and *lncNAT2$_{prom}$*::*GUS* expression changes antagonistically upon leaf maturation. In contrast to $_{prom}$*::GUS* expression dynamics in leaves, both *UGT73C5$_{prom}$*::*GUS* and *lncNAT1$_{prom}$*::*GUS* are simultaneously co-expressed in roots (Supplementary Figure 1). Completely developed leaves at 25 DAS shows only a marginal $_{prom}$*::GUS* activity for *UGT73C5*. *LncNAT2$_{prom}$*::*GUS* is faintly active while *UGT73C6* promoters again exhibit the previously mentioned patchy expression pattern. Furthermore, both *UGT73C6$_{prom}$::GUS* and *UG773C5$_{prom}$::GUS* do not show overlapping expression pattern with *lncNAT2$_{prom}$*::*GUS* in young leaves (Figure 8 & Supplementary Figure 1). These results provided an overview of the gene expression pattern for *lncNATs-UGT73C6, UGT73C6* and *UGT73C5*. Reporter activity analysis points out that expression of *UGT73C6, UGT73C5* and *lncNATs-UGT73C6* is spatio-temporally distinct and developmentally regulated.

In order to check whether expression pattern observed in the reporter constructs of *lncNATs-UGT73C6, UGT73C6* and *UGT73C5* is in accordance with the endogenous transcripts abundance, a time course assessment for the expression levels was carried out by quantitative real-time (qPCR) (Figure 8 C). Supplementing the results obtained from reporter GUS assays, expression analysis showed higher levels for *UGT73C5* whereas *UGT73C6* is poorly expressed compared to *lncNATs-UGT73C6* across the developmental stages analyzed in this study except at senescence (45 DAS). Interestingly, quantification of transcripts levels from whole plant showed

that expression dynamics of *lncNATs-UGT73C6* (presumably *lncNAT2*) and *UGT73C5* were positively correlated during progression of various growth stages of *A. thaliana* (Figure 8 C). Collectively, our analysis verified that the annotated *lncNATs-UGT73C6* possess distinct promoters and are expressed in a developmentally regulated manner in *A. thaliana.*
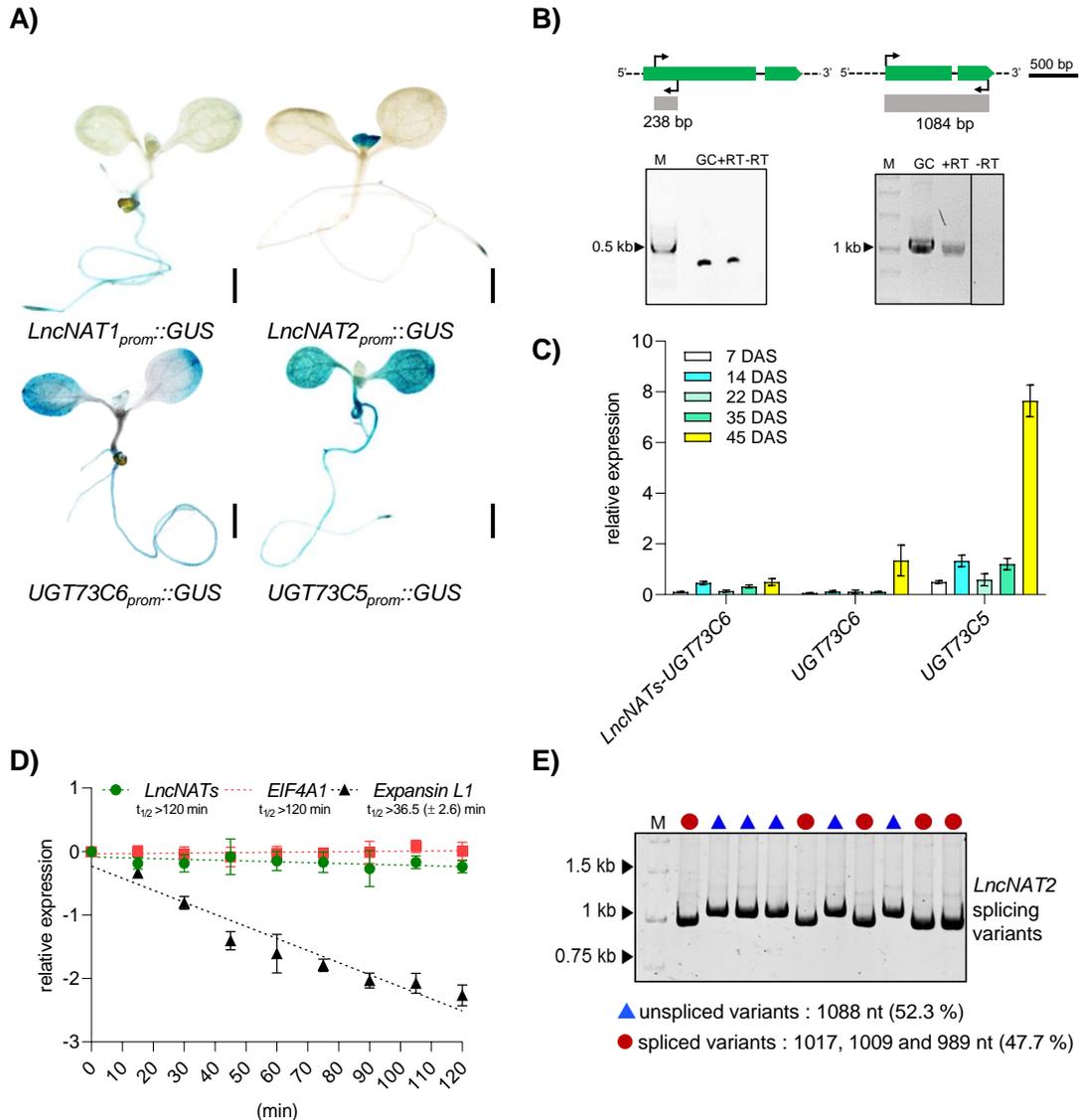


Figure 8: **Characteristics of *lncNATs-UGT73C6*.** A) Seedlings at 7 days after sowing (DAS) displaying *prom::GUS* expression pattern for *lncNAT1*, *lncNAT2, UGT73C6* and *UGT73C5*. Scale bar 5 mm. B) Primer scheme for the amplification of a fragment of *lncNAT1* and the complete *lncNAT2* (top panel). Arrow heads (black) and bars (grey) denote size of the amplicon. RT-PCR illustrating detection of *lncNAT1* and *lncNAT2* (bottom panel). M: Marker, GC: Genomic control, RT: reverse transcription. C) Time course expression analysis of *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* at different developmental stages. D) mRNA degradation curves for *lncNATs-UGT73C6* and control transcripts *EUKARYOTIC TRANSLATION INITIATION FACTOR 4A1* (*ELF4A1)* and *Expansin L1.* Half-life ($t_{1/2}$) is indicated. n=3, error bar represents standard deviation (±SD) in C& D. E) Colony-PCR amplification of *lncNAT2* variants. Size and percentage of clones carrying spliced and unspliced *lncNAT2* variants is indicated.

## 5.2 *LncNATs-UGT73C6* transcripts are stable and alternatively spliced

To gain further insights into what can additionally affect differences in the expression levels of *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* other than the promoter activity itself, we performed RNA stability assays in *Col-0.* The determination of half-life[3] for intron-less *UGT73C5* and *UGT73C6* mRNAs was not possible from the standard curve method (Ms. Susane Engelmann, personal communication). Both protein-coding transcripts behaves in same way and degrades rather quickly after the cordycepin treatment within less than ~20 minutes, an observation that seems to be in accordance with the reduced stability of intron-less mRNAs (Wang et al., 2007). Whereas *lncNATs-UGT73C6* were found to exhibit a half-life of more than 2 hours (Figure 8 D). Increased stability of *lncNATs-UGT73C6* may additionally contributes to overall abundance of *lncNATs-UGT73C6*. However, our data suggest that mRNA stability does not contribute for observed differences in the transcript abundances of *UGT73C6* and *UGT73C5.* Therefore, difference in the levels of *UGT73C6* and *UGT73C5* mRNAs in *A. thaliana* potentially occurs due to the differences in promoter activity.

Both *lncNATs-UGT73C6* encoding genes possess an intron in the 3' region (Figure 4). To further characterize *lncNATs-UGT73C6* transcripts, we analyzed if the intron in *lncNAT2* is subject to canonical splicing. RT-PCR product of full length *lncNAT2,* which was obtained by using gene specific primers in cDNA synthesis, was cloned and inserts present in individual colonies were analyzed using gene specific primers and verified by sequencing. Reported intron was retained in approximately 50% of the colonies while remaining 50% showed that *lncNAT2* RNA is spliced[4] (Figure 8 E). Furthermore, qPCR quantification of *lncNATs-UGT73C6* from nuclear and cytoplasmic RNA extractions indicated that *lncNATs-UGT73C6* are localized in the cytoplasm[5] (Ms. Susane Engelmann, personal communication). Collectively, these results confirmed that *lncNATs-UGT73C6* are transcribed from specific promoters, stable, alternatively spliced and transported to the cytosol, suggesting that they can act in *trans*.

## 5.3 Constitutively overexpressed *lncNAT1* increases rosette area

In order to study the role of *lncNAT1* in *A. thaliana* development, we generated transgenic lines overexpressing the full-length genomic *lncNAT1* sequence under the control of the strong

---

[3] RNA stability assays by using Cordycepin were performed by Ms. Susane Engelmann
[4] These experiments were part of master thesis pursued by Mr. Micheal Fritz.
[5] qPCR quantification of *lncNATs* from nuclear and cytoplasmic fraction was performed by Ms. Susane Engelmann

*Cauliflower mosaic virus* (*CaMV*) *35S* promoter in *Col-0* background. Overexpression of *lncNAT1* was confirmed by northern blot for the selection of independent overexpression lines (*lncNAT1ox*) for the phenotypic analysis (Supplementary Figure 2 A).

In a first experiment, 20 plants for each transgenic line were grown in long day conditions (16hour light/ 8 hour dark cycle) with controlled light (~100-120 $\mu mol \cdot m^{-2} \cdot s^{-1}$), temperature (21° C day/18° C night) and humidity (75%) including optimal irrigation and randomized distribution of the pots that harbors transgenic lines. Phenotypic observations at several developmental stages revealed that *lncNAT1* overexpression lines i.e. *#3.9*, *#4.2* and *#13.3*, *#15.4* develops rosettes that are bigger than control plants carrying the empty vector (Figure 9 A). Further quantification of total leaf area, from the digital images of complete rosette of individual plants, was performed by using Easy Leaf Area Program ® (Easlon and Bloom, 2014) (Supplementary Figure 3). Compared to plants transformed with the empty vector, *lncNAT1ox* lines *#3.9*, *#4.2*, *#13.3* and *#15.4* showed significantly increased rosette size at 25 DAS (Figure 9 B) or 35 DAS. Additionally, 20 plants each for 2 to 3 lines transformed with the empty vector and *Col-0* were grown together and quantified for leaf area. Comparison of leaf area showed no differences between *Col-0* and independent transgenic lines transformed with the empty vector. A comprehensive summary of three independent experiments showing phenotypic effects of *lncNAT1* overexpression are outlined in Table 2. Combined analysis from different experiments performed either in growth cabinets or in greenhouse showed an average increase of 16.28 (±5.1), 14.5 (±6.2), 9.35 (±0.35) and 9.71 (±3.2) % in rosette area at 25 DAS for *lncNAT1ox* lines *#3.9*, *#4.2*, *#13.3* and *#15.4.* Considering the average increase in rosette area in the overexpression lines of *lncNAT1* from all experiments (Table 2) a total increase of 12.45 % was observed with ±3.4 % standard deviation (SD). However, it is important to note that level of *lncNAT1* overexpression is not strongly correlated with the extent of increase in leaf area (Supplementary Figure 2 A, Figure 9 B and Table 2). The increase in rosette area observed due to *lncNAT1* overexpression prompted us to further examine the phenotypic effects of overexpression of *lncNAT2,* which is expressed in leaves.

**A)**
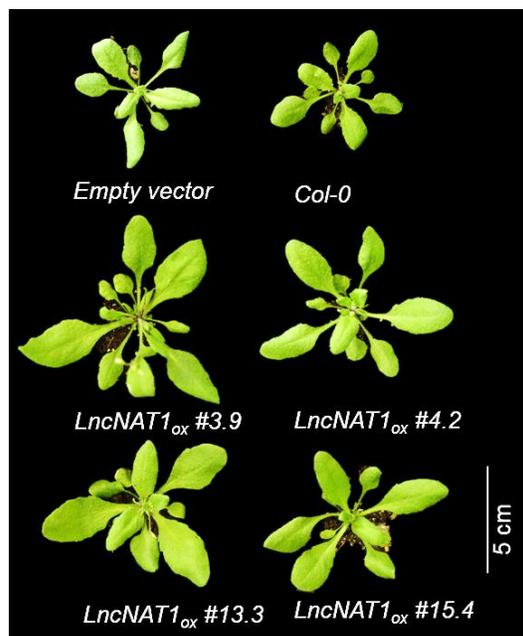
**B)**



Figure 9: ***LncNAT1* overexpression increases leaf area in *A. thaliana***. A) (Left) Complete rosette images of representative 25 days old wild type (*Col-0*) and transgenic homozygous T$_3$ plants transformed with the empty vector and overexpressing *lncNAT1* (lines viz: *#3.9*, *#4.2*, *#13.3* and *#15.4*) under the control of *35S* promoter. B) Box plots showing rosette area distribution in control and transgenic lines. Data in box plots shows minimum and maximum rosette area (whiskers) including outliers, first and third quartiles (boxes) and median values (vertical lines inside the boxes). Data shown are representative of 1 of minimum 3 independent experiments. Each group contains 20 individual plants that were grown on soil in long day conditions in greenhouse (~100-120 µmol·m$^{-2}$·s$^{-1}$ light intensity, temperature 21°C day/18°C night, 75% humidity and irrigated with a fixed volume of water once in a week). Asterisks indicate statistical significance compared to plants transformed with the empty vector. * <0.05, one-way ANOVA. n.s. – not significant.

## 5.4 Constitutively overexpressed *lncNAT2* increases rosette area

In a similar manner to the overexpression of *lncNAT1,* full-length genomic *lncNAT2* was overexpressed under the control of *CaMV 35S* promoter in *Col-0* background. *LncNAT2* overexpression was confirmed (Supplementary Figure 2 B) by qPCR over cDNA performed with gene specific primer for *lncNATs-UGT73C6* (Table 7). Independent T$_3$ homozygous *lncNAT2* overexpression lines (*lncNAT2$_{ox}$*) *#6.3*, *#9.3*, *#13.4* and *#24.3* were selected for studying the phenotypic effects of *lncNAT2* overexpression.

Quantification of rosette area from 25 DAS plants grown in the greenhouse under controlled conditions showed that all *lncNAT2$_{ox}$* lines developed significantly larger rosette area

compared to control plants transformed with the empty vector (Figure 10) and similar results were obtained for plants grown in growth cabinet at 35 DAS. Comparison of rosette area of $lncNAT2_{ox}$ plants with 25 days old $lncNAT1_{ox}$ plants that were grown in similar growth conditions in greenhouse (Figure 9) revealed that overexpression of either of $lncNATs$-$UGT73C6$ results in similar phenotypic effect i.e. increase in rosette size (Figures 9 and 10). Collective analysis of three independent experiments (Table 3) at 25 DAS showed that $lncNAT2_{ox}$ *#6.3, #9.3, #13.4* and *#24.3* exhibit an average 21.6 (±10.2) %, 29.74 (±7.0), 31.8 (±7) % and 18 (±9.2) % increase in rosette area. An average value of 25.2 (±5.6) % increase in rosette area was observed for *lncNAT2* overexpression lines from all experiments shown in Table 3. Comparison of the change in rosette area for various overexpression lines for *lncNAT1* and *lncNAT2* shows that phenotypic effects of *lncNATs-UGT73C6* are comparable (Table 2 and 3). However, the increase in rosette area was not proportionally correlated with levels of *lncNAT2* overexpression (Supplementary Figure 2 B) as well. Together, in repeatedly performed experiments overexpression of each of the *lncNATs-UGT73C6* consistently resulted in increased rosette area (Tables 2 and 3, Figures 9 and 10).

Apart from the representative experiments described above, a global overview of the several leaf area experiments, including additional sets of independent $lncNAT2_{ox}$ lines generated during the investigation, showing phenotypic effects due to overexpression of *lncNATs-UGT73C6* are listed in Table 4. Analysis of leaf area at early developmental stages i.e. 16 DAS did not show obvious phenotypic effects. Our examination revealed that phenotypic effects due to overexpression of *lncNATs-UGT73C6* becomes more apparent only during later phases of development i.e. after 25 DAS (Supplementary Figure 4). For this reason, leaf area quantifications were mostly performed at 25 DAS or 35 DAS to study the role of *lncNATs-UGT73C6.* However, we preferably restricted our quantitative analysis of rosette area to 25 DAS because at 35 DAS plants develop flowers and a number of superimposed small leaves that are part of rosette irrespective of the fact that phenotypic effect is stronger at the later time point (Supplementary Figure 4). Taken together these experiments reinforced specific involvement of *lncNAT2* in determination of final leaf size in *A. thaliana.* Consequently, further studies were particularly emphasized over the elucidation of the role of *lncNAT2* in development of *A. thaliana* leaves.

**A)**

**B)**



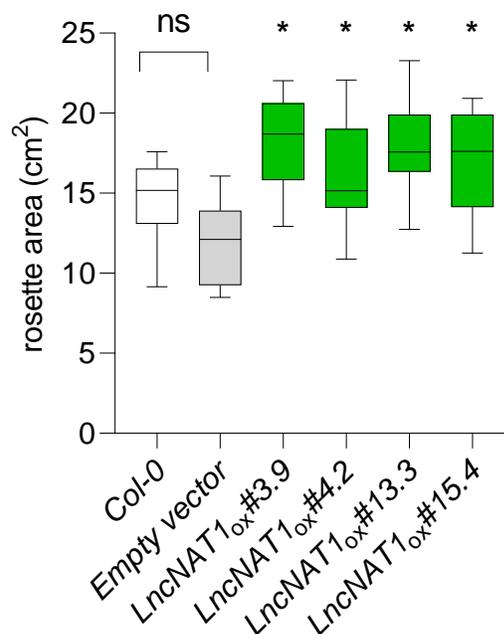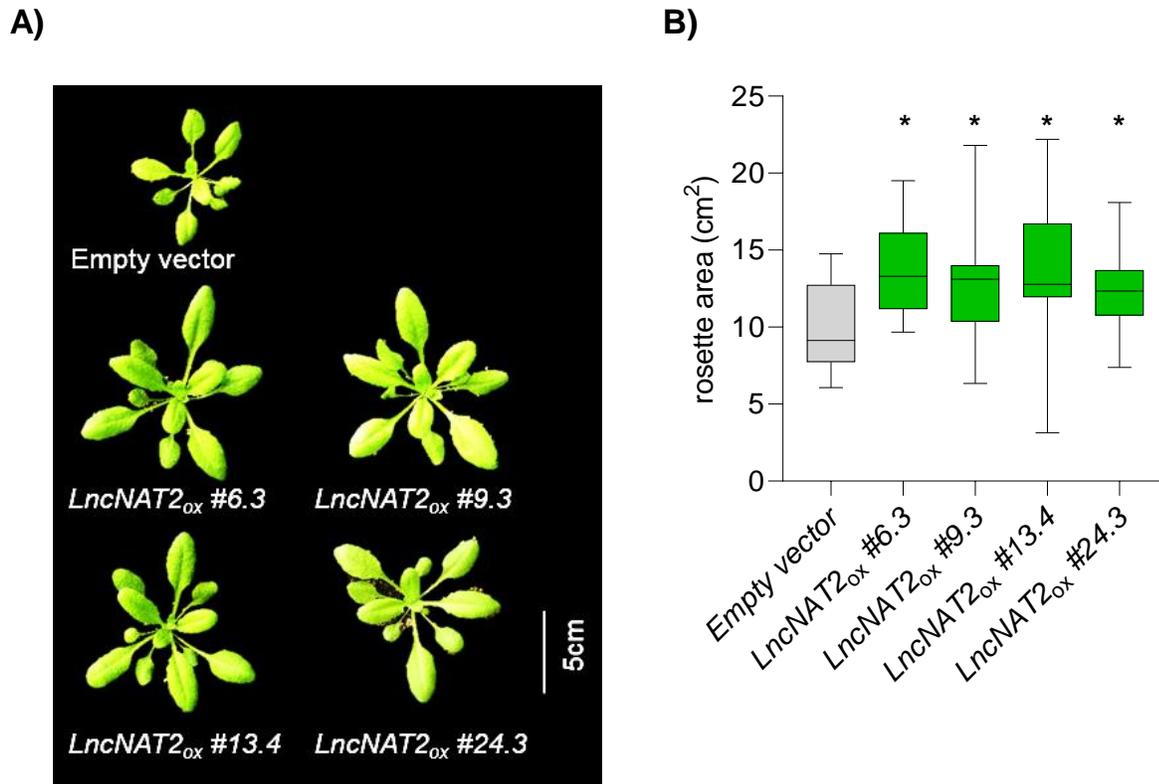Figure 10: ***LncNAT2* overexpression increases leaf area in *Arabidopsis*.** A) (Left) Complete rosette images of representative 25 days old transgenic plants transformed with the empty vector and overexpressing *lncNAT2* (independent $T_3$ homozygous *lncNAT2* overexpression lines viz: *#6.3, #9.3, #13.4* and *#24.3*). B) Box plots showing rosette area distribution in control and transgenic lines. Plant were grown on soil in greenhouse under long day conditions during summer months. Data shown are representative of 1 of minimum 3 independent experiments. Other details are same as for legend to Figure 10. Asterisks indicate statistical significance compared to plants transformed with the empty vector. n=20 plants per line, p * <0.05, one-way ANOVA.

## 5.5 Downregulation of *lncNATs-UGT73C6* in *trans* results in reduced rosette area

In order to study whether the downregulation of *lncNATs-UGT73C6* in *trans* affects leaf area, we generated transgenic lines expressing *artificial microRNAs* (*amiRNAs*) in *Col-0* background. Using *mir390* backbone, *amiRNAs* constructs were designed to downregulate *lncNATs-UGT73C6* post-transcriptionally in *trans*-specific manner according to the procedure reported earlier (Carbonell et al., 2014). Three *amiRNA* constructs, named as *amiRNA1*, *2* and *3*, were generated to produce perfect complementary guide sequences (Table 7) to target both *lncNAT1* and *2* transcripts at different locations (Figure 11 A) and expressed under the strong *35S* promoter. Northern blots using $^{32}$P $\gamma$-ATP labeled oligonucleotide probes showed that all three

*amiRNAs* were expressed and transcripts of all the constructs were precisely processed into 21nt long *amiRNAs* (Figure 11 A and Supplementary Figure 5) and that the highest detection levels was observed for the mature *amiRNA1* in comparison to *amiRNA2* and 3. qPCR expression analysis showed that the degree of downregulation of *lncNATs-UGT73C6* varies in the different constructs and lines. A maximum of ~50% downregulation of *lncNATs-UGT73C6* could be achieved only for a*miRNA1* construct followed by *amiRNA2* and *3* to a lesser degree (Figure 11 B).

Detailed phenotypic characterization of *amiRNAs* lines, in similar way to that of *lncNATs-UGT73C6* overexpression, was carried out in both growth cabinets and greenhouse conditions. Significantly smaller rosettes were observed in *amiRNA 1.3.3*, *amiRNA 2.1.1* and *amiRNA 3.3.2* transgenic lines compared to plants transformed with the empty vector (Figure 11 C). Transgenic lines in which *amiRNA* fails to downregulate *lncNATs-UGT73C6* did not show changes in rosette area (line *amiRNA 3.8.4*, Figure 11 B & D). *amiRNA1.3.3,* the line with the maximum downregulation of *lncNATs-UGT73C6,* showed strongest phenotype opposite to that of *lncNATs-UGT73C6* overexpression effects over rosette area. A summary of four independent experiments performed using *amiRNA* constructs for independent *amiRNA* lines is presented in Table 5. Analysis of three independent experiments performed in the greenhouse (1 to 3G, Table 5) showed the highest reduction in rosette area in line *amiRNA 1.3.3* (27.5 (±0.73) %), whereas lines *amiRNA 2.1.1* and *amiRNA 3.3.2* showed an average decrease of 18 (±7.1) and 18.8 (±11.6) %, respectively (Table 5). These results highlighted that downregulation of *lncNATs-UGT73C6* in *trans* leads to the phenotypic effect opposite to that of the *lncNATs-UGT73C6* overexpression. Based upon these results, we concluded that alteration in the transcript abundance of *lncNATs-UGT73C6* in *trans* modulates rosette area suggesting that *lncNATs-UGT73C6* plays role in determining leaf area in *A. thaliana*.
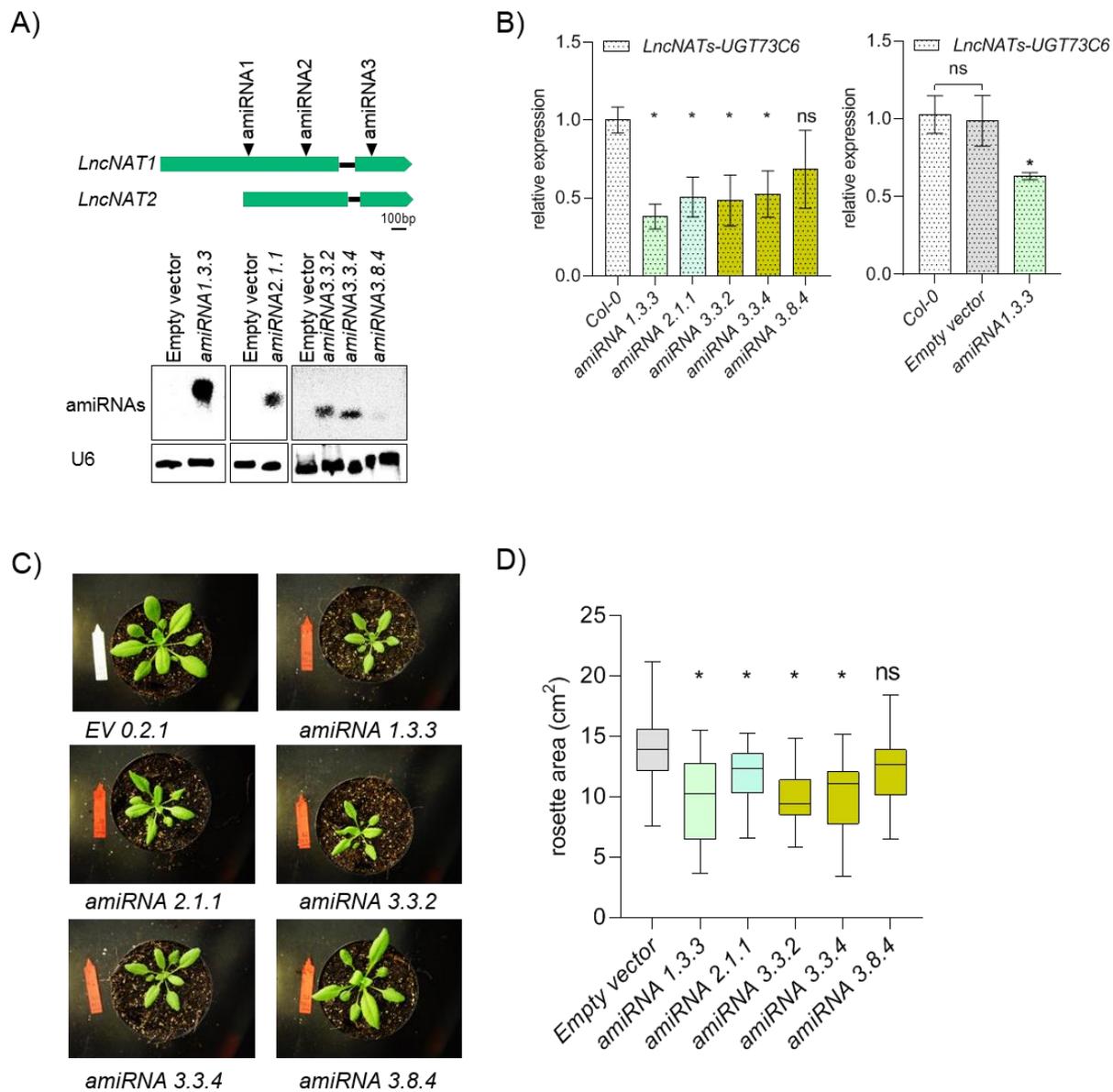
Figure 11: **Downregulation *lncNATs-UGT73C6* in *trans* decreases rosette area**. A) Scheme showing target site of *amiRNAs* over *lncNATs-UGT73C6* molecule (upper panel). Northern blot detection of *amiRNAs* in homozygous *A. thaliana* transgenic lines carrying *35S::amiRNA* constructs (bottom panel). $^{32}$P γ-ATP labeled oligonucleotide probes i.e. RR236, RR237, RR238 are used to detect *amiRNA1*, *2* and *3* respectively. First number in each transgenic line indicate respective *amiRNA* construct. *U6* (detected with oligo probe RR85) RNA is included as control. B) Downregulation of *lncNATs-UGT73C6* by amiRNAs constructs compared to *Col-0* (left) and an additional experiment showing downregulation of *lncNATs-UGT73C6* in *amiRNA 1.3.3* compared to *Col-0* and empty vector (right). n=3, *p <0.05, error bar represents standard deviation (±SD). C) Pictures of 25 DAS representative plants transformed with the empty vector and *amiRNA* lines showing effects of *lncNATs-UGT73C6* downregulation. D) Box plots showing quantification of rosette area in 25 DAS plants from representative experiment. n=20, * p<0.05, one-way ANOVA.

## 5.6  *LncNAT2* is a bona fide long noncoding RNA

*In-silico* prediction of *pORFs* shows lack of coding potential in *lncNAT2* which hinted that annotated 1084 nt long *lncNAT2* might function as a NAT-lncRNA. Previous studies in our laboratory (de-Vries, 2016) anticipated a maximum of 5 potential small *pORF*s in the unspliced variant of *lncNAT2 viz: pORF1, 2, 3, 3'* and *4.* Transient expression assays in *N. benthamiana* showed translation of *GFP*-tagged peptides 1, 2 and 3 when overexpressed under the *35S* promoter and peptide 3 showed specific nuclear localization (de-Vries, 2016) (Figure 7). Therefore, in order to address whether *lncNAT2* acts as bona fide NAT-lncRNA a two tier strategy was adopted.

First, transgenic plants carrying C-terminally tagged GFP fusions for *pORF1*, *2*, *3* and *4* were individually overexpressed in the context of *lncNAT2*[6] under the *35S* promoter for studying the phenotypic effects of its overexpression. Quantification of rosette area for three independent homozygous lines for each construct showed that except *lncNAT2$_{1-404}$ pORF3 GFP #8.4* none of the overexpression lines produces plants with bigger rosette area compared to *Col-0* (Figure 12). *pORF3* was additionally overexpressed without the *lncNAT2* upstream sequence but its overexpression was not able to induce an increase in rosette area. In summary, 5 out of six independent *pORF3* overexpression lines both with and without the *lncNAT2* upstream sequences did not show increase in rosette area (Figure 12). Why some of the *pORFs* overexpression lines i.e. *lncNAT2$_{1-404}$ pORF3 GFP #1.6, lncNAT2 pORF3 GFP #1.8, lncNAT2 pORF3 GFP #3.3* and *lncNAT2$_{1-742}$ pORF3 GFP #2.5* exhibited reduced rosette area was out of the scope of this thesis and therefore was not investigated further. Notwithstanding, this analysis supported our conclusion that overexpression of peptides encoded in *lncNAT2* sequence are not responsible for the increase in rosette area and that *lncNAT2* is a bona fide NAT-lncRNA.

---

[6] *LncNAT2* context: *pORFs* transgenic constructs having adjacent 5' upstream sequence of *lncNAT2* prior to the start sites of pORFs were termed as in the context of *lncNAT2* whereas *pORFs* only sequence within *lncNAT2* was termed as out of the context.
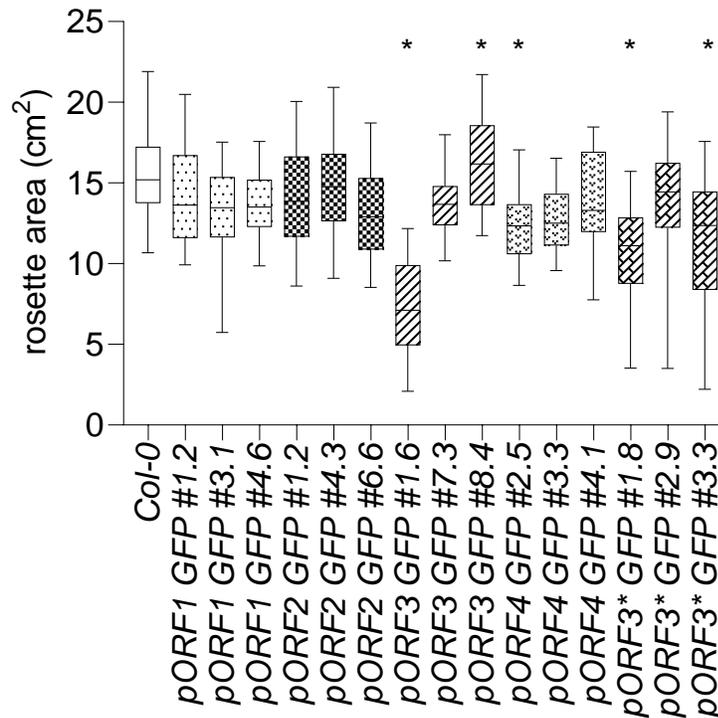
Figure 12: **Overexpression of peptides encoded by *lncNAT2* does not result in increased rosette area.** Box plot showing shows rosette area for 25 DAS plants for *Col-0* and independent *35S::pORF 1*, *2* and *3* transgenic lines. In the construct for *pORF1, 2, 3* and *4*, respective 5' upstream *lncNAT2* sequence i.e. 1-198, 1-288, 1-404 and 1-742 is included. In each construct the stop codon was deleted to allow the translation of fusion GFP. In transgenic lines, labelled as *lncNAT2 pORF3\* #1.8*, *#2.9* and *#3.3*, *pORF 3* is overexpressed as an individual ORF without upstream *lncNAT2* sequence. Combined data from 2 independent experiments, n=30, *p<0.05, one-way ANOVA.

Second, translation start sites of *pORF1*, *2, 3* and *3'* were site directed mutagenized (SDM) by changing ATG codon to CCC in WT *lncNAT2* sequence (Figure 13 A). Resulting mutated form of *lncNAT2* (hereafter termed as *Mut_lncNAT2*), was overexpressed in parallel to the WT *lncNAT2* under *35S* promoter in *Col-0* background. Additionally, unlike *pORF1* and *2*, *pORF3* can potentially encode a peptide of 103 amino acid long, a condition that would disqualify *lncNAT2* as lncRNA (Ben Amor et al., 2009) (Figure 7 C). Therefore, an additional mutant variant was generated in which only the translation start site of *pORF3* was mutated (*lncNAT2_mut_ORF3*) to CCC by SDM. Thus, in *Mut_lncNAT2* the possibility of peptide synthesis from all *pORFs* was eliminated while in *lncNAT2_mut_ORF3* other *pORFs* can still be translated. *In silico* structural analysis showed that mutation(s) minimally affects the overall *lncNAT2* structure. Also, the levels of overexpression were confirmed in all lines (Mr. M. Heidecker, personal communication). Phenotypic analysis of *Mut_lncNAT2* and *lncNAT2_mut_ORF3* was carried as described earlier

(section 5.3 and 5.4). Overexpression of WT *lncNAT2* or *Mut_lncNAT2* resulted in increased rosette area compared to *Col-0* or to the plants transformed with the empty vector (Figure 13 A). Three independent overexpression lines each for WT *lncNAT2* or *Mut_lncNAT2* consistently, confirming that the observed global increase in rosette is due to the RNA molecule and not due to the potential peptides.
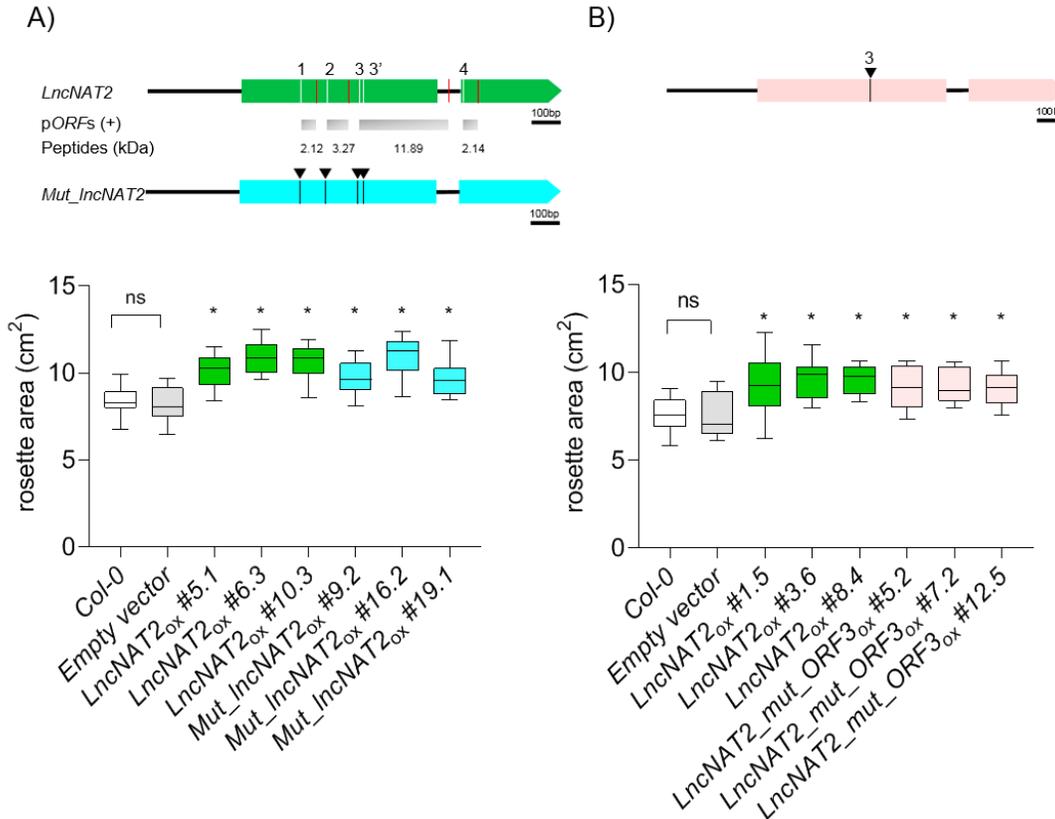


Figure 13: ***LncNAT2 is a bona fide long noncoding RNA***. A) Genomic organization of *lncNAT2* showing *pORF 1, 2, 3, 3' and 4*. White and red lines represent putative start and stop codon in wild type *lncNAT2*. Black arrows denote ATG to CCC mutation at the translation start site (ATG) for *pORF1, 2, 3 and 3'* in *Mut_lncNAT2* (top). Box plots showing rosette area from one representative experiment at 25 DAS for *Col-0*, empty vector, and overexpression lines of WT *lncNAT2* and *mut_lncNAT2* (bottom). n=18, *p<0.05, one-way ANOVA. B) Scheme of mutagenesis for *pORF3* (top). Data from 1 out of 2 independent experiments with similar results showing quantification of rosette area for *Col-0* and empty vector and overexpression lines of *lncNAT2_mut_ORF3* at 25 DAS. n=15, *p<0.05, one-way ANOVA.

Three independent overexpression lines of *lncNAT2_mut_ORF3* viz: *#5.2*, *#7.2*, and *#12.5* also showed the same phenotypic effects to that of the *lncNAT2* overexpression lines confirming that peptide encoded by *pORF3* cannot be the causative factor for increase in rosette area (Figure 13 B). It is to be noted that the comparison of mutant lines was always carried out with the freshly

generated $T_3$ transgenic lines overexpressing *lncNAT2* or transformed with the empty vector to exclude the effects of seeds aging and differences in growth conditions. Also, in each experiment *Col-0* seeds were used that were harvested at the time of the collection of $T_3$ seeds for different transgenic lines.

## 5.7  *LncNAT2* overexpression results in enlarged mesophyll cell size at the leaf bottom

*LncNAT2$_{prom}$*::*GUS* showed stronger activity in developing leaves and its expression decreases during leaf morphogenesis. Additionally, we found that number of leaves is not affected due to the overexpression of *lncNAT2* (Supplementary Figure 6). Therefore, development of larger rosettes due to the overexpression of *lncNAT2* could possibly occur as a consequence of alterations in sub-anatomical parameters such as the number and/or size of cells. We confirmed changes in rosette (Figure 13 A) and leaf (Figure 14 A) area in additional *lncNAT2$_{ox}$* lines and performed microscopic analysis.

The selection procedure of two $6^{th}$ leaves for the microscopic analysis is described in the following section. The plants were grown in the pots with soil under tightly controlled environment in growth cabinet. Soon after the appearance of the $6^{th}$ leaves all plants were marked. For the selection of 2 individual $6^{th}$ leaves from a pool of 15 plants, we first quantified rosette area 25 DAS for *lncNAT2$_{ox}$ #5.1, lncNAT2$_{ox}$ #6.3* and *lncNAT2$_{ox}$ #10.3* and *Col-0* to confirm that phenotypic effects are evident (Figure 14 A). Based upon similar median leaf area in different overexpression lines, out of *lncNAT2$_{ox}$ 6.3* and *lncNAT2$_{ox}$ 10.3, lncNAT2$_{ox}$ 10.3* was chosen for microscopy. At 35 days these plants were further subjected for quantification of rosette area (data not shown). Consequentially marked $6^{th}$ leaves from plants, which exhibited uniform and similar rosette area nearer to the respective median value in each group, were dissected at 35 days, photographed and quantified. Analysis of area of $6^{th}$ leaf from 8 individual plants showed that phenotypic effects due to *lncNAT2* overexpression also occurs locally at individual leaf level. $6^{th}$ leaf area in *lncNAT2$_{ox}$ 10.3* line was significantly bigger than *Col-0* (Figure 14 A). After verifying that area of individual leaf is also affected due to overexpression of *lncNAT2*, we, proceeded to select and dissect 2 individual leaves from two different plants for *Col-0* and *lncNAT2$_{ox}$ #10.3*. Subsequently microscopic analysis was performed according to a procedure reported earlier (Wang and Guo, 2015) (Supplementary Figure 7). Optical sectioning using differential interference contrast (DIC) mode in light microscope (Apotome 2®) was performed for unstained and completely transparent leaf segments from tip and bottom region of leaf to examine the effect of *lncNAT2* overexpression.

Analysis revealed that majority of adaxial epidermal cells at the bottom side of leaf showed typical puzzle-like morphology (Sapala et al., 2019) with no apparent changes in the dimensions of cells in *Col-0* and *lncNAT2ox #10.3* line. A few epidermal cells were extraordinarily bigger in size in *lncNAT2ox #10.3* compared to *Col-0* both at the tip and bottom (Figure 14 B). Estimation of total number of cells in a fixed area showed that epidermal and palisade mesophyll cells were significantly less in *lncNAT2ox 10.3* compared to *Col-0* at the bottom (Supplementary Figure 8, Table 6). The number of palisade mesophyll cells per image field is ~28% less in *lncNAT2ox 10.3* than the *Col-0* at bottom of leaf. While the size of palisade mesophyll cells was twice to that of *Col-0* in *lncNAT2ox #10.3* line in same region of leaf indicating that the change in number of cells is consequent to enlargement in area (Figure 14 B and Supplementary Figure 8). In contrast, with the exception of few bigger cells, majority of epidermal cells in *Col-0* and *lncNAT2ox #10.3* exhibited similar cell area at the bottom of the leaf suggesting that *lncNAT2* overexpression effects in epidermis are less stronger than in mesophyll cells. However the total numbers of all cell types, excluding guard cells, in bottom epidermis layer is significantly smaller in *lncNAT2ox 10.3* compared to *Col-0.*

Interestingly, examination of cellular changes at the tip of leaf due to the *lncNAT2* overexpression showed no apparent differences in size and number of either of the selected cell types between *Col-0* and *lncNAT2ox #10.3* (Figure 14 B and Supplementary Figure 8). Due to a large variation in the size of different cell sub-types in epidermis layer, an overall rough approximation for total number of cells in complete leaf using mesophyll cells was easier. The median area of mesophyll cells from both the tip and bottom was used for calculation of number of mesophyll cells in the entire leaf. Compared to *Col-0*, *lncNAT2ox #10.3* showed a total of ~16 % decrease in the number of mesophyll cells in the whole leaf (Table 6). However ~16% reduction in number of cells could not be strictly correlated with ~2 fold increase in area of cells as described above. Thus, our results suggest that *lncNAT2* induced effects are more localized at bottom region of leaf and are stronger in the mesophyll cells.

To examine if the overexpression effects of *Mut_lncNAT2* are similar to that of the overexpression of WT *lncNAT2*, we performed the same microscopic analysis. The phenotypic output due to the overexpression of *Mut_lncNAT2ox #16.2* is comparable to that of *lncNAT2ox #10.3* (Figure 14 A). Overexpression of *Mut_lncNAT2ox #16.2* also affected mesophyll cell area only at the leaf bottom similar to that of *lncNAT2ox #10.3* (Figure 14 B, left). Compared to *Col-0*, both *Mut_lncNAT2ox #16.2* and *lncNAT2ox #10.3* lines constantly exhibit bigger area for mesophyll cells

at the bottom of leaf suggesting that phenotypic changes occur due to the RNA and not by the peptides.
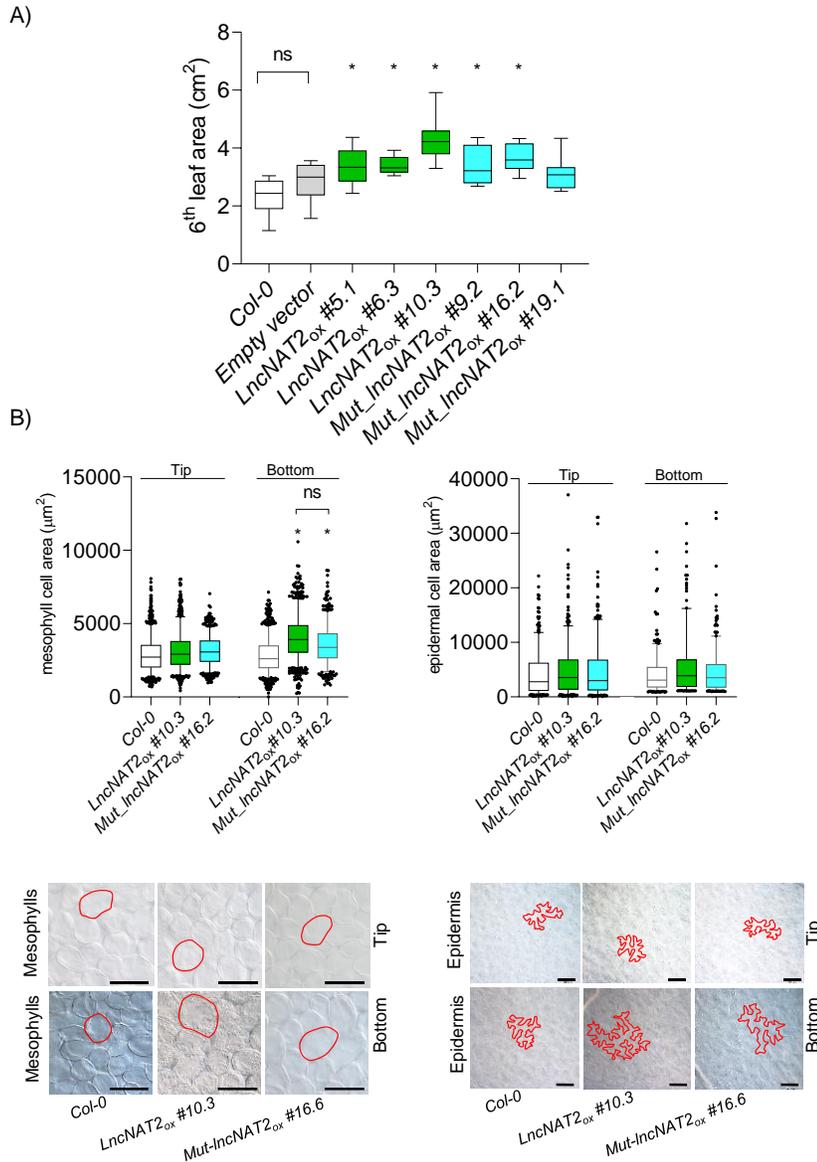


Figure 14: **Macroscopic and microscopic effects of the overexpression of *lncNAT2* and mutated *lncNAT2* (*Mut_lncNAT2*).** A) Quantification of 6th leaf area from 35 DAS *Col-0*, empty vector, and overexpression lines (indicated by number after name of genotype) of *lncNAT2* and *mut_lncNAT2*. n=8, * $p<0.05$ (one-way ANOVA). B) Box plot showing area of mesophyll (left) and epidermal (right) cells in tip and bottom regions of the leaf (upper panel). Representative pictures showing one highlighted mesophyll (left) and epidermal (right) cell in tip and bottom region of the leaf (lower panel). A total of 726 to 1321, and 599 to 979 mesophyll cells were used to determine cell area in tip and bottom regions, respectively. For epidermal cells, a total of 449 to 505 and 262 to 362 epidermal cells were used to quantify cell area in tip and bottom regions, respectively. Box plots shows 5-95 percentile and outliers are indicated. * $p<0.05$, one-way ANOVA. Scale bar 100 µM.

Epidermal cells were not affected though the presence of few exceptionally bigger cells observed again in *Mut_lncNAT2$_{ox}$ #16.2* compared to *Col-0* (Figure 14 B, right). The number of cells remained unaltered at the tip region of leaf in *Mut_lncNAT2 16.2* for both the mesophyll and epidermal cells while a similar effect was observed at bottom region of leaf (Supplementary Figure 8). Collectively, our analysis ratifies that the *lncNAT2* affects leaf size in *A. thaliana*, acting as a bona fide NAT-lncRNA.

## 5.8  Phenotypic effects of *UGT73C6* overexpression and downregulation

In order to analyze and independently reproduce reported results (Husar et al., 2011) for *UGT73C6,* we overexpressed the gene in a manner similar to *lncNATs-UGT73C6* overexpression. Levels of overexpression of four independent lines are shown in Figure 15 B. The representative images of plants (Figure 15 A) show that out of 3 overexpression lines, only *UGT73C6$_{ox}$ #9.5,* with 115 fold higher levels of *UGT73C6,* show skewed leaf morphology with reduced growth compared to *Col-0*. *UGT73C6$_{ox}$ #4.2* and *#14.2* plants with respective 30 and 41 fold higher levels of *UGT73C6* showed no obvious cabbage phenotype as reported earlier (Husar et al., 2011). Moreover, alterations in leaf morphology (Figure 15 A, *UGT73C6$_{ox}$ #5.1* and *#14.2*) could only be observed when the plants were grown at high light intensity (150 µmol·m$^{-2}$·s$^{-1}$ instead of 100-120 µmol·m$^{-2}$·s$^{-1}$ applied for phenotypic analysis of *lncNATs-UGT73C6*). Unlike *UGT73C6*, overexpression of *UGT73C5* consistently displays a typical BR deficiency phenotype (Figure 15 A) independent of the light intensity. Thus, in our analysis, *UGT73C6* overexpression does not lead to a clear distinguishable phenotype characteristic of BR deficiency (Gonzalez et al., 2010). These results provide a basis to argue that *UGT73C6* potentially does not play a primary role in maintaining BR homeostasis in *A. thaliana*.

On the other hand, to analyze the effects of *UGT73C6* downregulation, a previously described T-DNA insertion mutant line (SAIL_525_H07) (Jones et al., 2003) for *UGT73C6 (ugt73c6$_{ko}$)* was used. Exact location of the insertion in the mutant line *ugt73c6$_{ko}$* was confirmed by genotyping using specific primers that binds in the insert and in the *UGT73C6* sequence. Consequently, sequencing of the PCR fragment amplified over genomic *ugt73c6$_{ko}$* DNA template showed that the T-DNA is inserted at the 3' end of *UGT73C6* coding region, 1440 bp downstream of *UGT73C6* transcription start site (Figure 16 A). qPCR quantification over cDNA performed with a gene specific primer (Table 7) that binds upstream of the T-DNA insertion (Figure 16 B) showed that the expression of *UGT73C6* is significantly abolished in the mutant line (Figure 16 B).
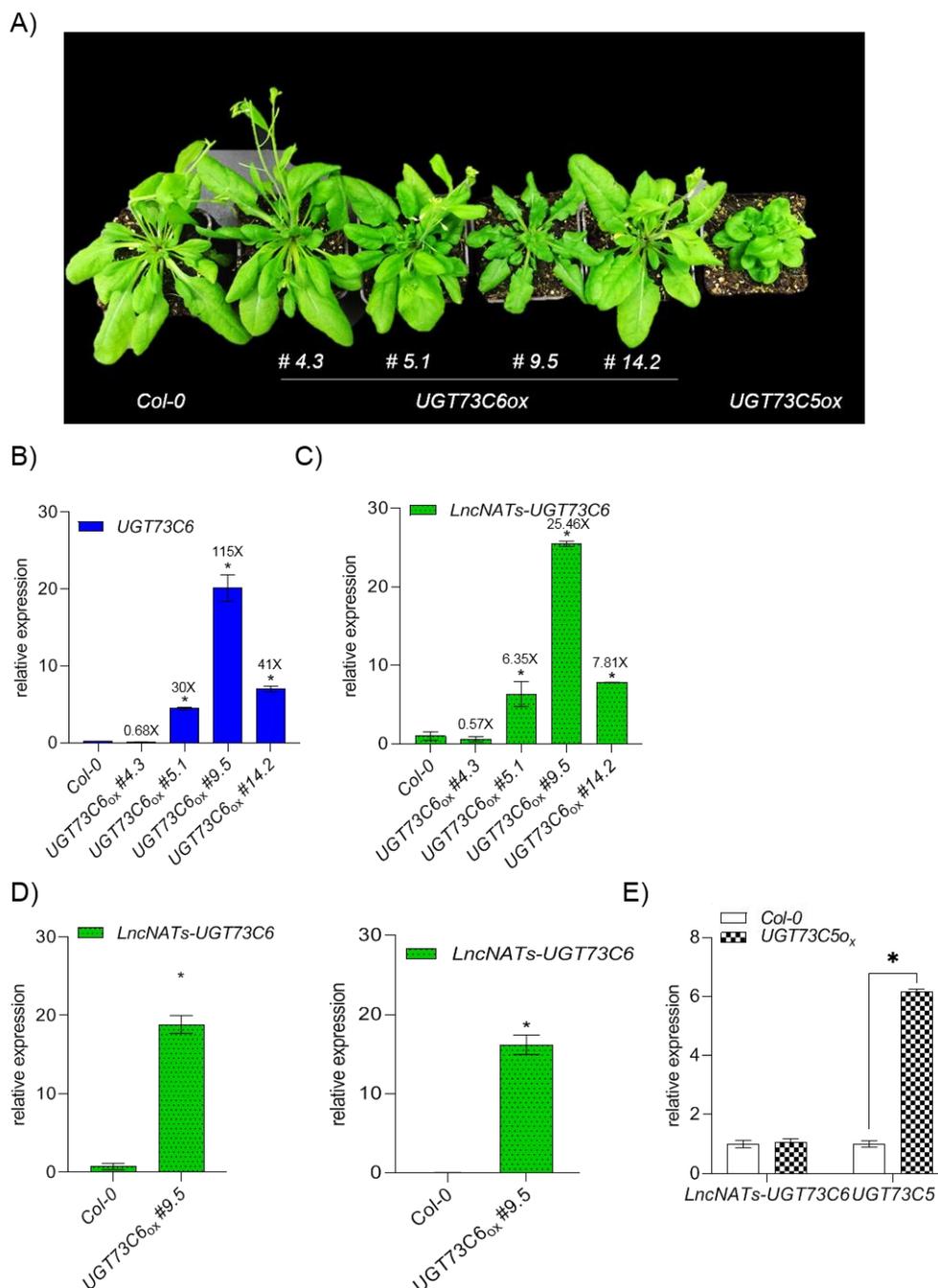
Figure 15: **Effect *of UGT73C6* overexpression over *lncNATs-UGT73C6* levels.** A) 4 week old representative plants overexpressing *UGT73C6* or *UGT73C5*. B & C) Levels of expression for *UGT73C6* and *lncNATs-UGT73C6* in WT and $UGT73C6_{ox}$ lines. Fold induction is indicated with respect to *Col-0*. D) qPCR quantification of *lncNATs-UGT73C6* in *WT* and $UGT73C6_{ox}$ *#9.5.* For cDNA synthesis, *lncNATs-UGT73C6* specific primer was used in (left) while cDNA synthesis was performed by employing primer that binds to *lncNATs-UGT73C6* transcripts originating specifically from the complementary strand of transgenic $UGT73C6_{ox}$ copy (right). $UGT73C6_{ox}$ constructs carry a linker sequence at 5' end from the vector. E) Relative expression levels of *lncNATs-UGT73C6* and *UGT73C5* in *Col-0* and $UGT73C5_{ox}$. 10 DAS seedlings used in all qPCR quantifications. ±SD, * $p < 0.05$. n=3, Student's t-test.
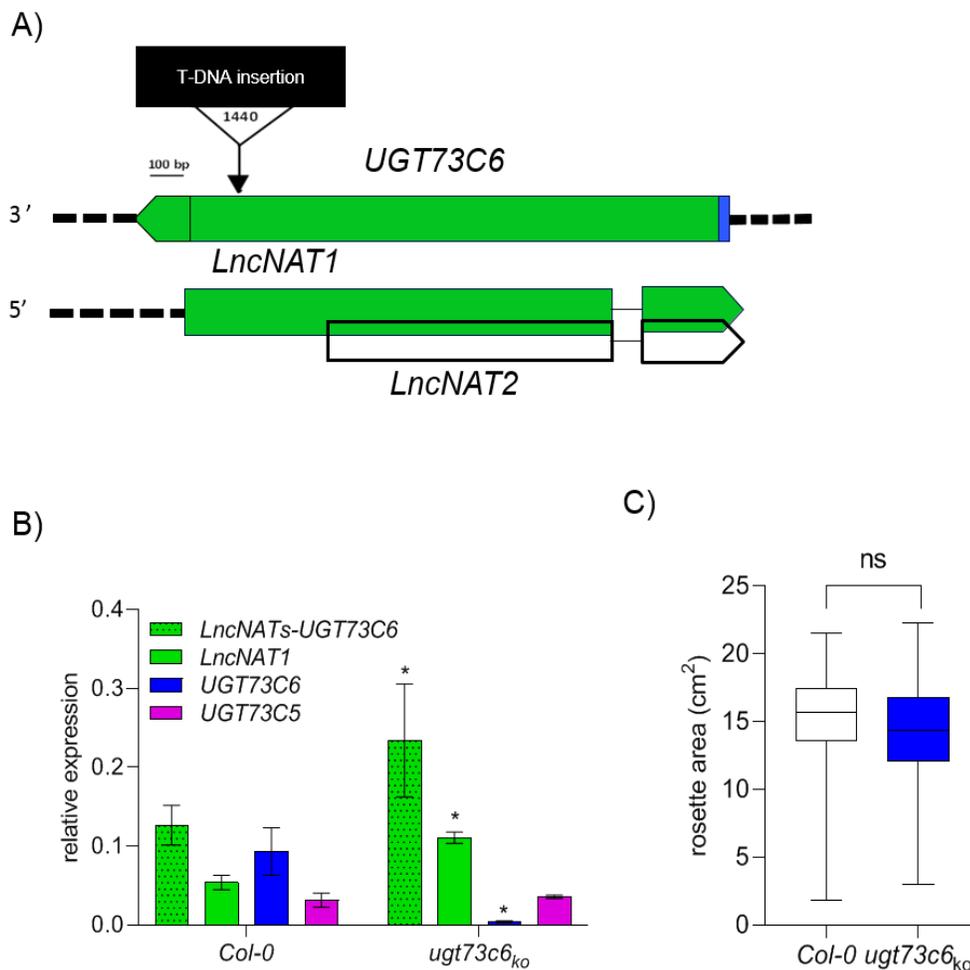
Figure 16: **T-DNA insertion in *ugt73c6ko* does not affect rosette area and expression of *lncNATs-UGT73C6*.**
A) Genomic organization of T-DNA insertion line *ugt73c6ko*. B) qPCR showing *UGT73C6* transcript levels in *Col-0* and *ugt73c6ko*. Depicted levels shown are average of 5 biological replicates. ** $p<0.001$, error bars represent ±SD, (student t-test). C) Box plot from 1 out of 3 independent experiments with similar results showing rosette area of 25 DAS *Col-0* and *ugt73c6ko* plants. n=30, $p>0.05$, One-way ANOVA.

Subsequent quantification of rosette area at 25 DAS displayed that *ugt73c6ko* and *Col-0* shows similar size (Figure 16 C). Collectively, these experiments clearly exhibited that rosette size is not affected due to absence of *UGT73C6* and only slight morphological changes takes place by high levels of *UGT73C6* overexpression.

## 5.9 Effects of alteration in *lncNATs-UGT73C6* expression over *UGT73C6* and *vice-versa*

Previously it has been shown that the reported NAT-lncRNAs in *A. thaliana vis-à-vis* regulate the expression of associated sense genes transcriptionally or post- transcriptionally via different mechanisms (Magistri et al., 2012; Ariel et al., 2015; Marchese et al., 2017). Among several possible means of gene regulation by *lncNATs-UGT73C6*, silencing of sense gene expression by nat-si-RNAs, which are generated due to DICER activity over complementary double-stranded sense and antisense RNA transcripts. Based upon previous reports (Borsani et al., 2005; Held et al., 2008), it was highly anticipated that NAT-lncRNAs can serve as precursors for nat-si-RNAs production because two complementary endogenous RNA molecules can form dsRNA leading to the generation of siRNAs due to DICER activity that is followed by RISC cleavage. Consequently, we postulated whether *lncNATs-UGT73C6* forms a locus oriented regulatory circuit to downregulate not only *UGT73C6* but also other closely related *UGT73C* family members due to high sequence similarity. Previous investigations showed that transient co-expression of *lncNATs-UGT73C6* with *UGT73C* family members in *N. benthamiana* can induce their downregulation (de-Vries, 2014). In this host, *lncNATs-UGT73C6* significantly downregulate *UGT73C6* and *UGT73C5* in addition to other *UGT73C* sub-family members when co-expressed (de-Vries, 2014). (Figure 6).

In order to check whether the above observation hold true for *A. thaliana*, we analyzed the levels of *UGT73C6* and its closest homolog *UGT73C5* in previously mentioned *lncNATs-UGT73C6* overexpression lines and downregulation lines. Independent overexpression lines for each of *lncNAT1* and *2* showed an unexpected elevation in the endogenous levels of *UGT73C6* whereas *UGT73C5* remains unaltered (Figure 17). In contrast, the downregulation of *lncNATs-UGT73C6* up to a maximum extent of ~53% to ~35% in *amiRNA* lines showed no detectable differences ($p > 0.05$) in levels of *UGT73C6* and *UGT73C5* compared to samples corresponding to plants transformed with the empty vector (Figure 17). Moreover, preliminary data indicated that alteration in *lncNATs-UGT73C6* quantity did not affect transcript abundance of other *UGT73C* sub-family members.
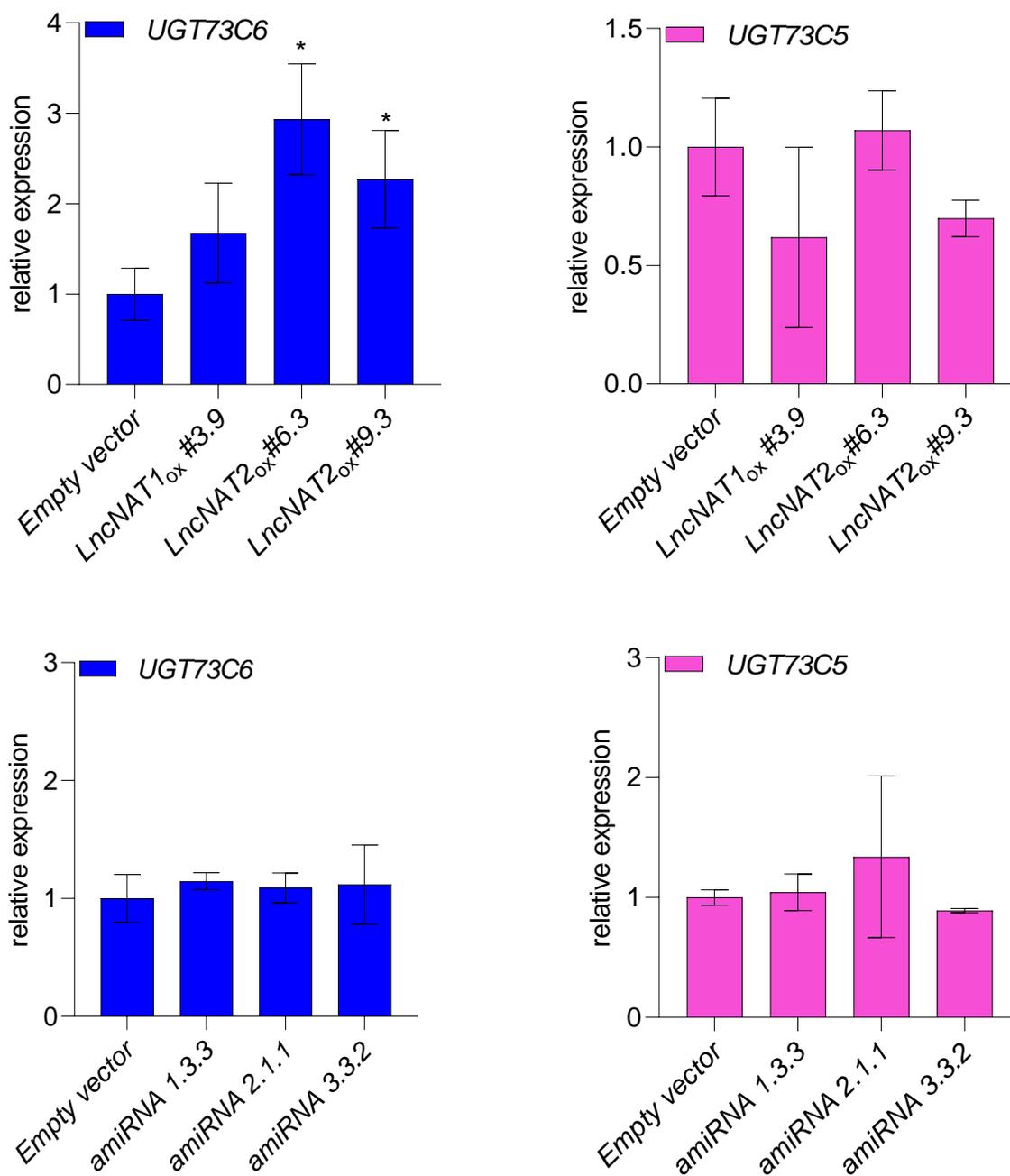
Figure 17: **Effects of alterations in *lncNATs-UGT73C6* expression levels over *UGT73C6* or *UGT73C5* in *trans*.** qRT-PCR expression analysis of *UGT73C6* (left) and *UGT73C5* (right) in transgenic 14 DAS *A. thaliana* plants overexpressing (upper panel) or downregulating (bottom panel) *lncNATs-UGT73C6*. Results from 1 representative experiment out of 3 with 3 biological replicates. Error bars indicate standard deviation (±SD), * $p < 0.05$, Student's t-test.

In addition, we argued if *UGT73C6* overexpression affects expression of endogenous *lncNATs-UGT73C6*. qPCR quantification over cDNA synthesized using gene specific primers showed a concomitant increase in levels of *lncNATs-UGT73C6* in the transgenic lines overexpressing *UGT73C6* (Figure 15 C). Extent of upsurge in *lncNATs-UGT73C6* abundance was comparable to the magnitude of *UGT73C6* overexpression in different lines. A maximum of 115 fold increase in expression of *UGT73C6* in *UGT73C6_{ox} #9.5* line induced ~25 fold increase in *lncNATs-UGT73C6* compared to ~6 and ~8 fold elevation in low overexpression lines *UGT73C6_{ox} #5.1* and *#14.2* that express respective 30 and 41 fold higher *UGT73C6* than *Col-0* (Figure 15 C & B). To differentiate whether the observed increase in *lncNATs-UGT73C6* arises from endogenous promoter activity or from the potential effects of *35S::UGT73C6* copy, cDNA synthesis was accomplished in *Col-0* and *UGT73C6_{ox} #9.5* line using primers (Table 7) that either binds to the 3' end of *lncNATs-UGT73C6* or to the transcripts containing the 28 nt sequence corresponding to the linker region from pB7WG2 binary vector in *35S::UGT73C6* (Supplementary Figure 9 A). Data shown in Figure 15 D confirms that elevated levels of *lncNATs-UGT73C6* in the *UGT73C6_{ox} #9.5* arise from the *35S* transgenic allele of *UGT73C6* and that endogenous *lncNATs-UGT73C6* promoter activity does not contribute to the induction in levels of *lncNATs-UGT73C6* upon *UGT73C6* overexpression. Further sequencing of PCR product for *lncNATs-UGT73C6* over cDNA synthesized with the oligonucleotide binding the sequence complementary to the linker region showed that accumulated *lncNATs-UGT73C6* does not undergo splicing unlike endogenous *lncNATs-UGT73C6* in *Col-0*[7] suggesting that they are the RDR products generated over *UGT73C6* mRNA. In contrast, *UGT73C5* overexpression does not affects the expression of *lncNATs-UGT73C6* (Figure 15 E) suggesting that observed affects are specific to *UGT73C6* although both genes are sister pairs (Figure 3). Our results indicate that potential RNA dependent RNA-polymerase (RDR) activity upon *UGT73C6* overexpression might be a reason for the elevation in levels of unspliced transcript corresponding to full length *lncNAT1*.Therefore, we concluded that overexpression of *UGT73C6* does not affects the endogenous levels of *lncNATs-UGT73C6*.

## 5.10 *LncNATs-UGT73C6* do not function via BR pathway

UGT73C6 and UGT73C5 are shown to deactivate brassinosteroids by 23-O-glycosylation (Husar, 2011;Poppenberger, 2005). To further examine the prospect of a BR driven anticipated gene regulatory loop amongst *UGT73C6*, *UGT73C5* and *lncNATs-UGT73C6,* we evaluated the

---

[7] These experiments were independently verified by Mr. M. Heidecker as part of his Master thesis

effect of most active BR, epi-brassinolide (epi-BL) in *Col-0* according to reported procedure (Husar et al., 2011). Detailed time course expression analysis by qPCR over a 24 hour period showed that control genes respond quickly soon after epi-BL treatment. 1 µM of epi-BL treatment induced consistent downregulation for chosen control BR biosynthetic genes viz: *BRASSINOSTEROID-6-OXIDASE 2* (*BR6ox2*) and *CYTOCHROME P450 90B1* (*DWF4*) post 0.5 hours (Supplementry Figure 10 B). Control BR catabolic gene *PHYB ACTIVATION TAGGED SUPPRESSOR 1* (*BAS1*) was induced as early as 1 hour (Figure 18 A). Conversely, *lncNATs-UGT73C6* remained largely unresponsive throughout the assay points except for a marginal but significant induction in transcript levels only at 2 and 3 hour after treatment (Figure 18 B).  Except for a slight induction of *UGT73C6* at 3 hour post treatment, both *UGT73C6* and *UGT73C5* were found to be largely unaffected by epi-BL treatment (Figure 18 C and D) in agreement with previously published reports (Husar et al., 2011).

To further support above findings that *lncNATs-UGT73C6* role is probably independent of BR pathway, we analysed expression levels of *ROT3* and *EXP8* in the transgenic lines in which *lncNATs-UGT73C6* are overexpressed or downregulated. ROT3 is required in late steps in BR biosynthesis whereas *EXP8* is a BR signalling gene and both genes are reported to play role in leaf development (Tsuge et al., 1996; Kim et al., 2005; Marowa et al., 2016). $UGT73C5_{ox}$ line has been included as postive control and shows significant alteration in expression levels (compared to *Col-0*) not only for *ROT3* and *EXP8* but also for two additional BR metabolic genes viz: *BR6ox2* and *CONSTITUTIVE PHOTOMORPHOGENIC DWARF* (*CPD*) (Figure 19 A) confirming that *UGT73C5* overexpression alters BR levels in planta as proposed previously (Poppenberger et al., 2005). Two independent *lncNAT2* overexpression lines, which show incresed rosette area (Table 3), exhibited no changes in *ROT3* and *EXP8* levels in comparision to *Col-0*. Also *amiRNA 1.3.3* downregulation line did not exhibit a clear antagonistic expression pattern, with respect to $lncNAT2_{ox}$ lines, for *ROT3* or *EXP8* (Figure 19 B). To further exclude that *lncNATs-UGT73C6* function is independent of BR signalling, measurements for etiolated hypocotyl growth were perfomed according to established procedure (Espinosa-Ruiz et al., 2015) using *Col-0, $lncNAT2_{ox}$ 6.3* and *amiRNA 1.3.3* lines. As expected, hypocotyl length was significantly shorter for $UGT73C5_{ox}$ and *cpd* insertional mutant (kindly provided by Prof. M. Quint) used as control in contrast to *Col-0* while $lncNAT2_{ox}$ *#6.3* and *amiRNA 1.3.3* showed no phenotypic effect (Figure 19 C).
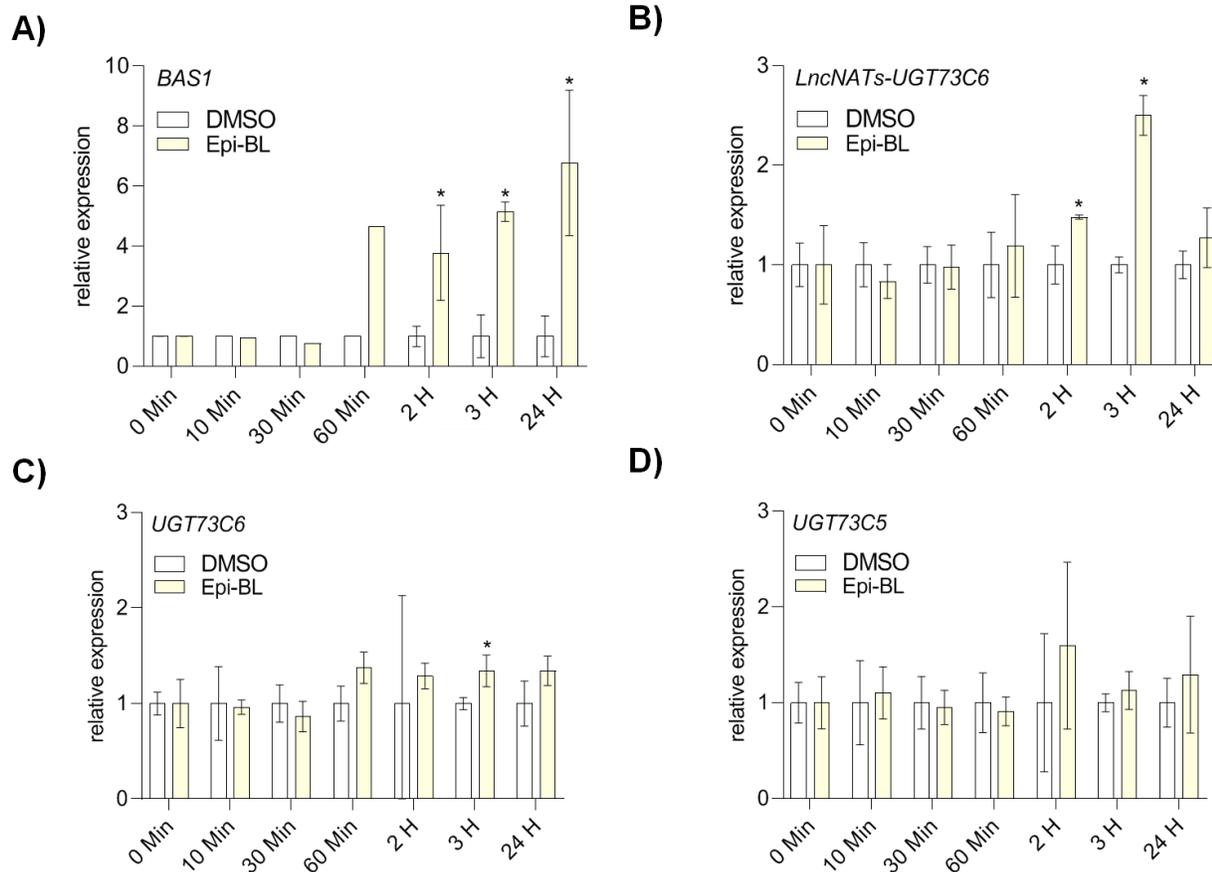
Figure 18: *LncNATs-UGT73C6* **expression is not induced by epi-BL.** 7 DAS seedlings were treated either with DMSO as control or 1µM epi-BL. qPCR expression analysis shows quantification of transcripts for control BR catabolic gene, *BAS1* (A), *lncNATs-UGT73C6* (B), *UGT73C6* (C) and *UGT73C5* (D) at different assays points. Data shown are combined from 3 independent experiments with 9 biological replicates in B, C and D. Values shown in A at time points viz; 10, 30 and 60 min are average from 2 biological replicates from two independent experiments while at time point 0 H, 2 H, 3 H and 24 H details are same as stated for B to D. Error bars are ±SD, * p<0.05, Student's t-test.

**A)**



**B)**



**C)**
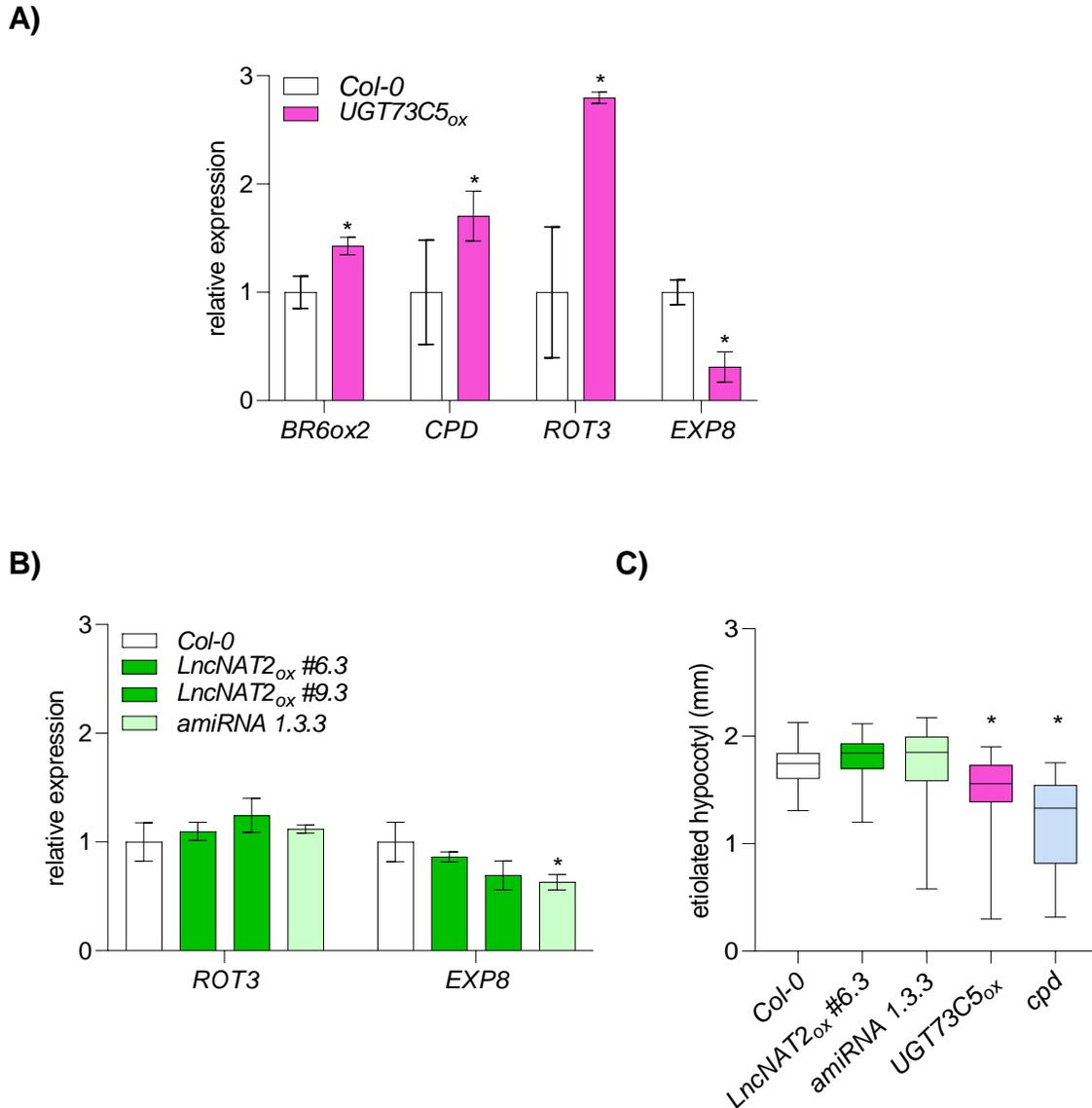


Figure 19: **Alterations in *lncNATs-UGT73C6* expression levels neither affects BR biosynthetic and signaling genes nor results in BR deficiency phenotype**. BR metabolic genes expression levels quantified in *Col-0* and *UGT73C5ox* (A) as a control to assess the effects of *lncNATs-UGT73C6* overexpression or downregulation over BR biosynthetic gene, *ROT3*, or over the signaling gene, *EXP8* (B). Experiment showing the length of etiolated hypocotyls for dark grown seedlings for indicated genotypes. *UGT73C5ox* and T-DNA insertion mutant for *CONSTITUTIVE PHOTOMORPHOGENIC DWARF* (*CPD*) are shown as control (C). (A to B) Error bars are ±SD, * $p < 0.05$. n=3, Student's t-test. C) n=40 seedlings, * $p < 0.05$, (one-way ANOVA).

Taken together we conclude that mentioned phenotypic effects on rosette area (Sections 5.3, 5.4 and 5.5 ) are not a consequence of BR driven regulatory loop employing *lncNATs-UGT73C6*.

## 5.11 Insights into the effect of genomic context over the expression of *lncNATs-UGT73C6, UGT73C6* and *UGT73C5*

Function of lncRNAs in *cis* or *trans* has been shown to be affected by the location of genomic locus (Kung et al., 2013; Han et al., 2018). Our results that were obtained after ectopic overexpression in *A. thaliana* highlighted role of *lncNATs-UGT73C6* in *trans*. It has been shown previously that specific genomic location can play crucial role in activation or repression of target genes by lncRNAs (Shechner et al., 2015). For instance in case of a novel *NUCLEAR ENRICHED ABUNDANT TRANSCRIPT 1* (*NEAT1*) lncRNA, the act of transcription at a particular genomic location itself was sufficient for its biological function related to paraspeckle (ribonucleoprotein bodies) formation (Mao et al., 2011). Observations from previous studies also prompted us to analyze the effect of proximity of *UGT73C6* with *UGT73C5* at the *UGT73C* locus to get an insight into the effect of locus over the expression *lncNATs-UGT73C6* both in and out of its genomic context. Moreover, such analysis can further help to understand the potential *cis* and/or *trans* roles of *lncNATs-UGT73C6*. Consequently, in order to evaluate the possible effect of genomic context over the expression profiles of *lncNATs-UGT73C6, UGT73C6* and neighboring *UGT73C5* we employed toxin response assays using a type B trichothecene mycotoxin viz: Deoxynivalenol (DON). Expression of *UGT73C6* and *UGT73C5* have been previously reported to be induced by the DON (Husar et al., 2011). To assess as how DON treatment affect induction and if position of locus plays role in the transcription o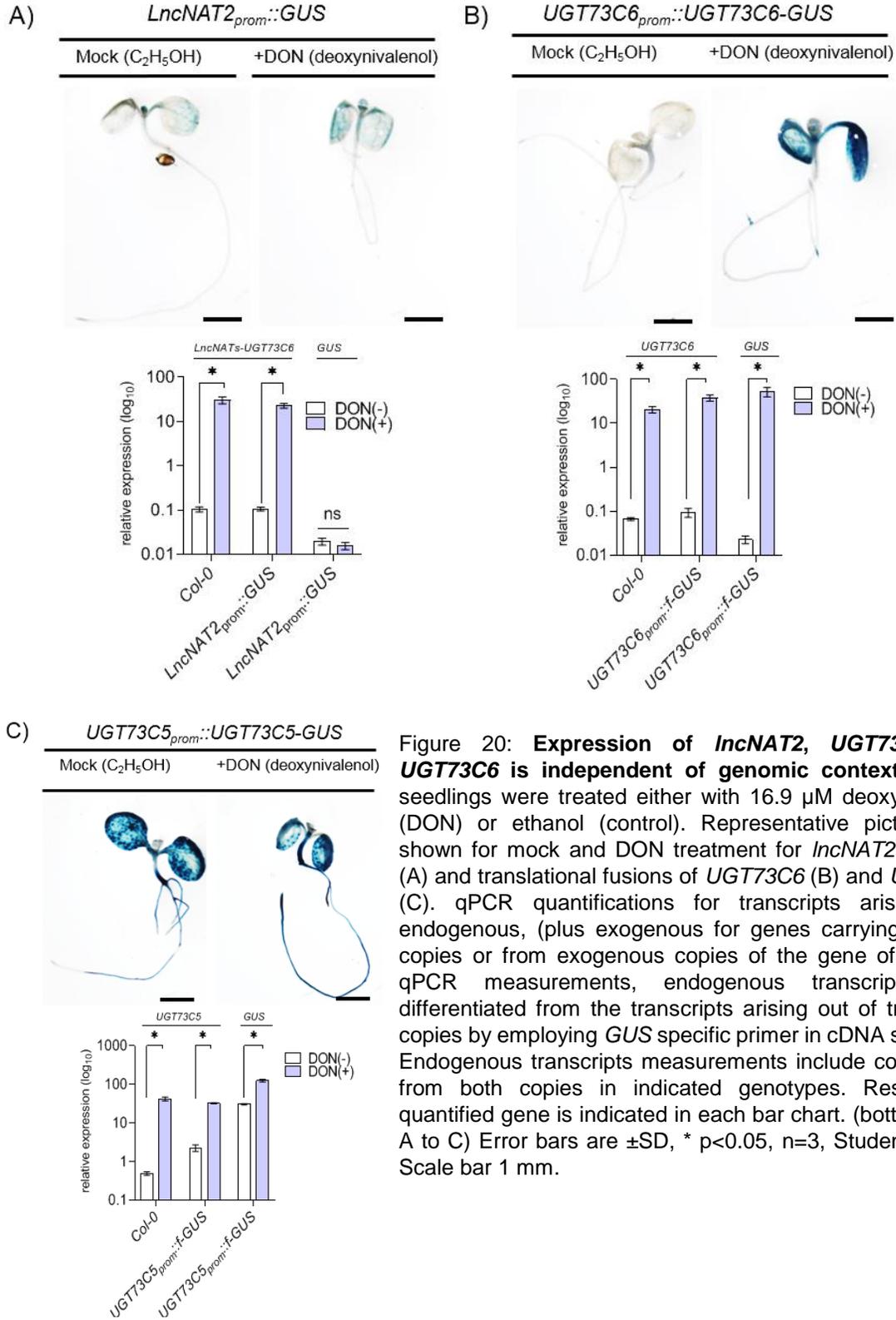f above genes, we employed reporter *GUS* and translational fusion lines. For *lncNAT1* and *lncNAT2, prom*::*GUS* lines were used while *UGT73C6prom::UGT73C6-GUS,* and *UGT73C5prom::UGT73C5-GUS* (kindly provided by Prof. B. Poppenberger) translational fusion lines were selected for *UGT73C6* and *UGT73C5*. 7 days old seedlings grown in liquid ½ MS media were exposed to either 16.9 µm of DON dissolved in ethanol or to the same volume of ethanol as control for 4 hours. Qualitative assay showed that GUS staining for *lncNAT2prom*::*GUS* remained apparently similar for DON treated and ethanol treated control samples (Figure 20 A). Also, *lncNAT1prom*::*GUS* activity showed no differences for GUS staining between treated and untreated samples (Supplementary Figure 11 B). *UGT73C6prom::f-GUS* line showed enhanced GUS staining upon DON treatment (Figure 20 B). However no noticeable apparent changes were observed in treated versus untreated samples for *UGT73C5prom::f-GUS* line (Figure 20 C). Different primers were used in cDNA synthesis for the further quantification by qPCR to discriminate whether the resulting transcripts originate from the endogenous gene in case of *lncNATs-UGT73C6* or from the endogenous plus the additional copy in the case of *UGT73C6* and

*UGT73C5* (gene specific primer) or from the randomly inserted transgenic alleles (*GUS*-specific primer) (Supplementary Figure 9 B). qPCR analysis involving quantification with gene specific primer showed that the transcription of *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* genes is strongly induced by DON treatment in *Col-0* and respective *GUS* fusion lines (Figure 20 A, B & C). In contrast, measurements of *GUS* specific transcripts showed no changes in expression levels for *lncNAT2$_{prom}$::GUS* after DON treatment (Figure 20 B)*. UGT73C6$_{prom}$::f-GUS* showed strongest elevation both in *UGT73C6* and *GUS* transcripts, that were close to two fold compared to DON treated *Col-0* seedlings*,* implying that the promoters of *UGT73C6* responds in similar manner to DON treatment in and out of the genomic context (Figure 20 B). On the other hand *UGT73C5* levels after DON treatment were nearly similar in *Col-0* and *UGT73C5$_{prom}$::f-GUS.* However, the observed higher levels of *GUS* transcripts for DON treated $_{prom}$::f-GUS* lines did not perfectly match with the quantified levels of *UGT73C6* and *UGT73C5* that were quantified using gene specific primers. These data suggest that, similar to *UGT73C6*, expression of *UGT73C5* is also independent of the genomic context although the levels of induction seems to have an upper limit for *UGT73C5.* Furthermore, the potential stabilizing effects of *GUS* sequence over the unstable *UGT73C5* or *UGT73C6* mRNAs seems to contribute to the increased expression levels of both genes (Figure 20 B & C). Considering the ethanol treated samples, it is important to take into account that the basal expression levels of *UGT73C5* in $_{prom}$::f-GUS* line are close to double in basal conditions compared to ethanol treated *Col-0* (Figure 20 C). Conversely, presence of an extra copy of *UGT73C6* in *UGT73C6$_{prom}$::f-GUS* does not result in expected doubling of the expression levels (Figure 20 B). On the other hand, unlike *UGT73C5,* DON treatment induces full induction of *UGT73C6* independent of the genomic context for both copies. Comparable expression levels after DON treatment in *Col-0* and *UGT73C5$_{prom}$::f-GUS* suggest that the endogenous copy of *UGT73C5* is fully induced while the out of the context copy is not. This observation highlights that *UGT73C5*, which doesn't have an associated NAT-lncRNA, responds differently to DON treatment when in and out *UGT73C* locus.

Out of the genomic context *lncNAT2$_{prom}$::GUS* allele showed no changes in *GUS* mRNA levels suggesting that elevation in levels of *lncNATs-UGT73C6* in *Col-0* or in *lncNAT2$_{prom}$::GUS* after DON treatment may not be dependent over the promoter activity of *lncNAT2* (Figure 20 A). Collectively, our analysis suggest that out of the context allele of *lncNAT2* can be transcribed independent of the locus as described earlier (section 5.1, Figure 8 A and Supplementary Figure 1) but remains to be unaffected by DON treatment (Figure 20 A). Unlike *lncNAT2$_{prom}$::GUS*, a small scale induction in the levels of *GUS* transcripts was observed for *lncNAT1$_{prom}$::GUS* after DON

treatment (Supplementary Figure 11 B).



A) *LncNAT2~prom~::GUS*

B) *UGT73C6~prom~::UGT73C6-GUS*

C) *UGT73C5~prom~::UGT73C5-GUS*

Figure 20: **Expression of *lncNAT2, UGT73C5* and *UGT73C6* is independent of genomic context.** 7 DAS seedlings were treated either with 16.9 µM deoxynivalenol (DON) or ethanol (control). Representative pictures are shown for mock and DON treatment for *lncNAT2~prom~::GUS* (A) and translational fusions of *UGT73C6* (B) and *UGT73C5* (C). qPCR quantifications for transcripts arising from endogenous, (plus exogenous for genes carrying fusions) copies or from exogenous copies of the gene of *GUS*. In qPCR measurements, endogenous transcripts were differentiated from the transcripts arising out of transgenic copies by employing *GUS* specific primer in cDNA synthesis. Endogenous transcripts measurements include contribution from both copies in indicated genotypes. Respectively quantified gene is indicated in each bar chart. (bottom panel A to C) Error bars are ±SD, * p<0.05, n=3, Student's t-test. Scale bar 1 mm.

To further discriminate whether accumulated *lncNATs-UGT73C6* transcripts are of *lncNAT1* or *lncNAT2* or from both, we performed RT-PCR for full-length endogenous *lncNATs-UGT73C6* transcripts in DON and ethanol treated samples. Results showed that a major portion of accumulated endogenous *lncNATs-UGT73C6* transcripts after DON treatment corresponds to the size of unspliced *lncNAT1* molecule indicating that high levels of endogenous *lncNATs-UGT73C6* are mainly due to *lncNAT1* (Ms. Susane Engelmann, personal communication). These data suggest that observed high levels of *lncNATs-UGT73C6* in DON treated samples could be due to potential RDR activity as described earlier for the overexpression of *UGT73C6* (Figure 15 C and D).

## 5.12 *LncNATs-UGT73C6* downregulation in *cis* causes pleotropic defects at early developmental stages

As mentioned earlier that *lncNATs-UGT73C6* is fully complementary to *UGT73C6* and, also, part of the promoter sequence upstream of annotated *lncNAT2* start site lies within the coding sequence of this gene. Due to this fact a strong constraint is imposed for the elucidation of the role of *lncNATs-UGT73C6* particularly in *cis* unlike the analysis of *lncNATs-UGT73C6* in *trans* by *amiRNAs* as shown in Figure 11. We were inspired by the following observations from *ugt73c6$_{ko}$* mutant line viz: 1) T-DNA insertion in 3' region of *UGT73C6* significantly abolished *UGT73C6* expression (Figure 16 B), 2) rosette area remains un-affected in *ugt73c6$_{ko}$* compared to *Col-0* (Figure 16 C), 3) *lncNAT2* is expression is not reduced in *ugt73c6$_{ko}$* (Figure 16 B) and 4) *UGT73C5* expression also remained unaffected (Figure 16 B) suggesting that insertion of T-DNA does not affect the closely related gene. Consecutively, to study of the role of *lncNATs-UGT73C6* in *cis* we employed CRISPR/Cas9 gene-editing approach in *A. thaliana* (Aparicio-Prat et al., 2015; Tsutsui and Higashiyama, 2017). A double *CRISPR* guide RNA approach was adopted to knockout *lncNATs-UGT73C6* at different regions by different combinations of primers (Figure 21 A and Table 7). $T_1$ positive transformants were selected based upon the fluorescence of seeds due to GFP expression driven by a seed-coat-specific promoter and successively the identity of $T_2$ and $T_3$ homozygous knockout lines was confirmed by genotyping and sequencing (Supplementary Figure 12). For complete and partial deletion of *lncNATs-UGT73C6, Cr_lncnats_*13.3 and *Cr_lncnats #18.2* lines respectively were selected in the *Col-0* background for further analysis. In *Cr_lncnats #13.3* 1750 bp of target sequence, starting at 45 bp downstream of annotated *lncNAT1* transcription start site until 183 bp downstream of annotated 3' end of *lncNATs-UGT73C6,* was

deleted leading to removal of *lncNAT1* and *2* and *UGT73C6* sequences. While *Cr_lncnats #18.2* line harbors a 603 bp long deletion at the 3' end of *lncNATs-UGT73C6* (Figure 21 A).
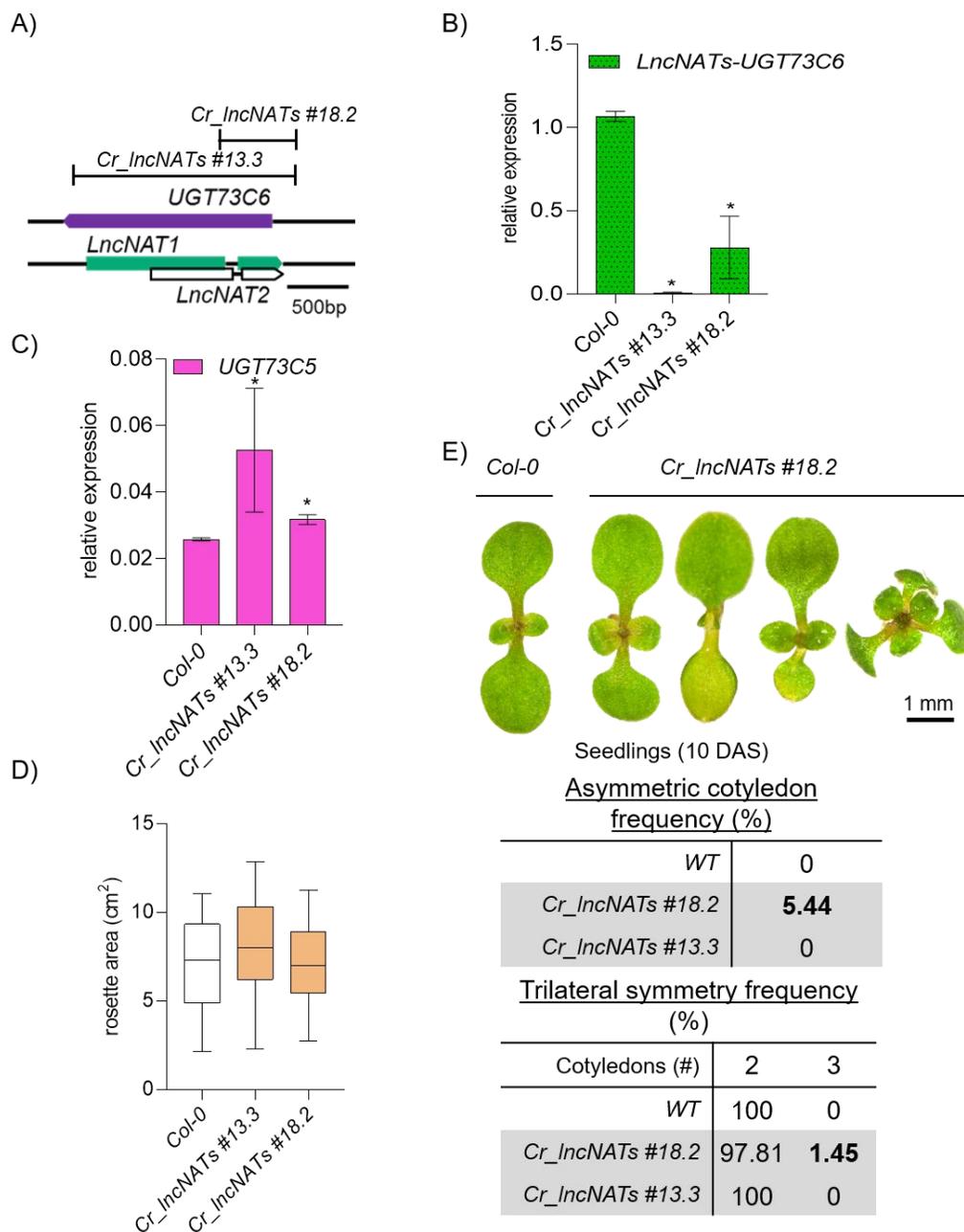


Figure 21: **Effects of *lncNATs-UGT73C6* knockout in *cis.*** A) Schematic showing CRISPR/Cas9 partial and full deletion of *lncNATs-UGT73C6* in *Col-0* background. B and C) Expression levels of *lncNATs-UGT73C6* and *UGT73C5* in knockout lines compared to *Col-0*. Error bars are ±SD, *p<0.05, n=3, Student's t-test. D) One out of three independent experiments showing quantified rosette area for 25 DAS plants grown in growth cabinet. n-20, p>0.05, one-way ANOVA. E) 10 days old *A. thaliana* seedlings showing effects of *lncNATs-UGT73C6* knockout in *Cr_lncNATs #18.2* line compared to *Col-0* (top panel).

The frequencies of asymmetric and trilateral symmetry of cotyledons in *Cr_lncNATs #18.2* and *Col-0* is shown (bottom panel).

qPCR data shows complete knockout of *lncNATs-UGT73C6* in *Cr_lncnats #13.3* while 70% knockdown was achieved in *Cr_lncnats #18.2* line that potentially expresses the first 509 nt of the long *lncNATs-UGT73C6* with a truncated 3' end (Figure 21 B). Additionally, small effect over the *UGT73C5* expression was observed (Figure 21 C). Surprisingly, in contrast to effects of downregulation of *lncNATs-UGT73C6* in *trans* (Figure 11 D), measurements of leaf area at 25D under controlled conditions showed plants with no changes in rosette leaf area when *lncNATs-UGT73C6* are fully or partially knocked out in *cis* compared to *Col-0* (Figure 21 D). Careful phenotypic observation of different developmental stages showed that knockdown of *lncNATs-UGT73C6* in *cis* causes pleiotropic effects leading to early developmental defects in a small part of the population (Figure 21 E). 10 to 14 days old seedlings of CRISPR/Cas9 knockout lines for *lncNATs-UGT73C6* showed that the shape of cotyledons turns asymmetric in subpopulation for *Cr_lncnats_*18.2 lines compared to *Col-0* (Figure 21 E). In addition to above effects, *Cr_lncnats #18.2* line showed seedlings with trilateral symmetry i.e. 3 cotyledons (Figure 21 E). The frequency of asymmetric cotyledons and trilateral symmetry shown in Figure 21 E. These data suggest that *lncNATs-UGT73C6* could play a role during early developmental stages.

## 5.13 *LncNAT2* knockout in *cis* partially affects *GRFs* expression

As mentioned before, changes in *lncNATs-UGT73C6* expression does not affect expression levels of *UGT73C6* or closely related *UGT73C* family members in *trans* (Figure17). Also, downregulation of *UGT73C6* in *ugt73c6$_{ko}$* mutant showed that rosette area is not affected (Figure 16 C) while at the same time promoter elements downstream of T-DNA insertion in *ugt73c6$_{ko}$* are sufficient enough to drive the transcription of *lncNATs-UGT73C6* to similar or slightly higher levels than that of *Col-0* (Figure 16 B). These data provided a strong line of evidence that *lncNATs-UGT73C6* mediated phenotypic effects are independent of *UGT73C6* and that *lncNATs-UGT73C6* do not establish a locus specific feedback regulatory loop for *UGT73C6*, *UGT73C5* and other *UGT73C* family members unlike as observed in previous transient transformation assays using *N. benthamiana* (de-Vries, 2014).

Our analysis showed that *lncNAT2$_{prom}$::GUS* activity is stronger in nascent young leaves and that tends to confine at the bottom region of leaf blade as the leaf matures (Figure 23). Interestingly, *lncNAT2$_{prom}$::GUS* expression pattern is similar to that of the *GRFs* (Debernardi et al., 2012). *GRF1* to *9* are expressed in *A. thaliana* and in association with the products of *GIF1, 2* and *3,* form regulatory complex at protein level to modulate various aspects of organ growth by

actively participating in cell division and differentiation processes (Omidbakhshfard et al., 2015) which helps in determination of the final organ size. Interestingly, *mir396a* and *b* are expressed in the leaf tip and due to their complementarity with *GRFs,* they direct cleavage of *GRFs* mRNA resulting in an antagonistic expression patterns at the bottom and tip of the leaf (Debernardi et al., 2012). Furthermore, *in silico* analysis showed that a 20 nt long stretch of RNA sequence in *lncNATs-UGT73C6* molecule near the 5' end of *lncNAT2* is partially complementary to *miR390* and forms a small bulge at the 11th nucleotide position that is required for endonucleolytic cleavage of target substrate mRNAs that is guided by *miR396* (Figure 23). Early developmental defects, similar to knockout of *lncNATs-UGT73C6* in *cis* (Figure 21 E), have been previously reported for the components of *GRF-mir396* regulatory module (Kanei et al., 2012). Based upon these observations we postulated that *lncNATs-UGT73C6* can affect the magnitude of expression of *GRFs* by acting as target mimic of *miR396*.
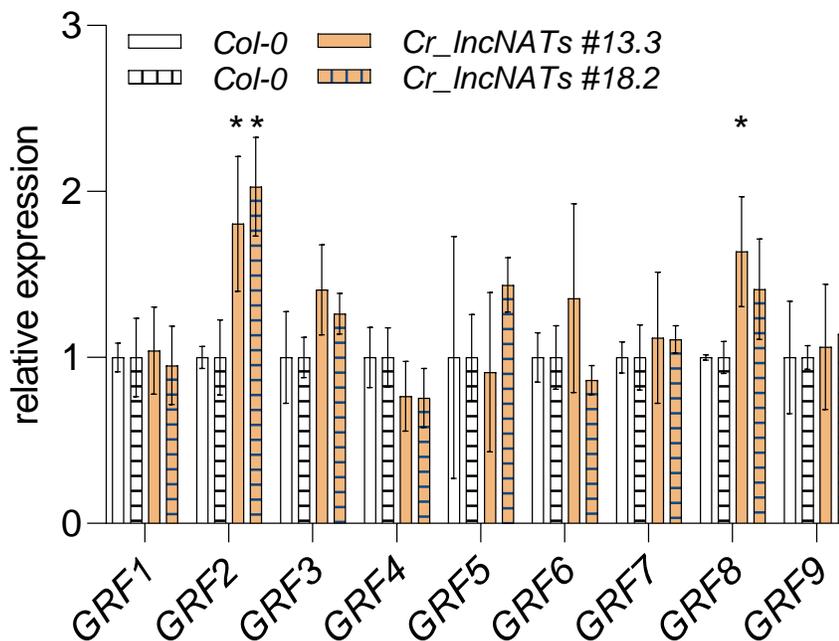


Figure 22: **Effects of alterations in *lncNATs-UGT73C6* expression levels over *GRFs* expression**. Abundance of *GRFs* in *Col-0* and CRISPR/Cas9 knockout lines of *lncNATs-UGT73C6.* Data are average of 3 biological replicates. Error bars are ±SD, * p<0.05, Student t-test.

Therefore, to prospect further into the potential underlining mechanism of *lncNATs-UGT73C6*, we sought to evaluate the effect of alteration in *lncNATs-UGT73C6* levels over *GRFs* abundance *in planta*. qPCR expression analysis was carried out in selected overexpression lines of *lncNAT2* and CRISPR/Cas9 knockout lines of *lncNATs-UGT73C6* using cotyledons or whole

seedlings. A preliminary screening of effects of *lncNAT2* overexpression showed a slight increase in *GRF4* and *GRF9* levels while *GRF 1* remained unaffected and *GRF 2*, *3*, and *5* were partially reduced upon overexpression of *lncNAT2* compared to *Col-0*. *GRF 4* and *9* are the most relevant targets of *mir396* regulatory module (Rodriguez et al., 2010). Unexpectedly, on the other hand, Figure 22 shows the effect over *GRFs* expression upon knockout of *lncNATs-UGT73C6* in *cis*. *GRF2, 3* and *8* showed the tendency of upregulation while *GRF4* is slightly downregulated in *lncNATs-UGT73C6* knockout lines compared to *Col-0*. Other *GRFs* were not affected due to knockout of *lncNATs-UGT73C6*. Notwithstanding, further *in vivo* and *in vitro* studies are required to validate if the *lncNATs-UGT73C6* acts as a target mimic of *miR396*. Also, dissection of young leaves into bottom and tip might facilitate to better resolve the expression analysis because *GRFs* are expressed in the bottom of leaves (Debernardi et al., 2012).

# 6 Discussion and future prospects

## 6.1 Expression of *lncNATs-UGT73C6* in *A. thaliana*

*LncNATs-UGT73C6* were originally annotated in publicly available databases TAIR (*Arabidopsis* information resource) and ARAPORT (Cheng et al., 2017) from RNA seq expression profiling studies as the two potential non-protein coding transcripts that originate from the opposite DNA strand of *UGT73C6*. As a firsthand attempt to investigate the functional relevance of both *lncNAT1* and *lncNAT2*, we sought to detect the *lncNATs-UGT73C6* transcripts in *A. thaliana*. TAIR reported apparently similar 3' end for both the transcripts. Therefore, using primers that binds to the annotated 3' ends in cDNA synthesis reaction, *lncNAT2* could be detected by RT-PCR using primers that amplifies full length annotated *lncNAT2* (Figure 8 B). However, we could not detect reported *lncNAT1* using similar strategy (data not shown) indicating that *lncNAT1* transcription start site could be further downstream of the reported 5' end. Subsequently walking RT-PCR method was adopted involving adjoining primers combinations at the different locations downstream of annotated 5' end. Using a forward primer annealing 108 nt downstream of the annotated 5' end in combination with the reverse primer, that bind upstream of *lncNAT2* start site, we were able to amplify fragment of *lncNAT1* (Figure 8 B). Such discrepancies with regard to databases could happen because of the 'reference annotation based transcript assembly' procedures in whole transcriptomic approaches for e.g. RNA-seq which often relies over the annotation of short sub-sequence of a cDNA sequence, termed as expressed sequence tags (EST), with already existing reference genome. The number of reeds, their alignment, and other factors may limits the resolution strength of RNA-seq as far as precise annotation of the exact ends of the transcripts is concerned (Ozsolak and Milos, 2011). Nevertheless, unlike *lncNAT2*, detected *lncNAT1* was found to be shorter than the annotated version. These studies were further supplemented by 'circular 5' rapid amplification of cDNA ends' (RACE) and conventional 3' RACE for *lncNAT1* and *lncNAT2*. RACE analysis showed that *lncNATs-UGT73C6* carry characteristic RNA Pol II polyadenylation mark, (poly-(A) tail), and slightly differ with reported ends of *lncNATs-UGT73C6* (Triller, 2019) indicating that in addition to alternative transcription initiation, potential leaky termination similar to reports in yeast (Candelli et al., 2018) may also be a source for variations in 3' ends of *lncNATs-UGT73C6* transcripts. Such genomic DNA elements and mechanistic features of transcriptional apparatus are in fact viewed as one of the evolutionary means for generation of multiple transcripts isoforms as reported in case of animal studies (Reyes and Huber, 2018). On the other hand alternative transcription initiation may simply be due to the

transcription initiation by the polymerase from non-optimal start sites and consequential transcripts may not carry functional or adaptive significance (Xu et al., 2019). This observation again reinforces debate over 'choice or noise' conventions of lncRNAs (Louro et al., 2009). In contrast, biologically relevant *COOLAIR* lncRNA exist as a pool of alternatively spliced isoforms of varied length and additionally the transcripts terminates either early or extends further into the *FLC* promoter region suggesting that loose regulation of transcription termination for lncRNAs might be a general phenomenon (Marquardt et al., 2014). Additionally, we observed that ~50% alternative splicing of *lncNATs-UGT73C6* also contributes to overall diversity for the *lncNATs-UGT73C6* transcripts. It might be noted that, despite presence of conserved splice sites, *lncNATs-UGT73C6* are poorly spliced resulting in roughly equal proportions of spliced and unspliced transcripts (Figure 8 E). Genome wide studies from animal sources pointed out that splicing of lncRNAs is generally inefficient and is proportional to the strength of 5' splice site (5'ss) altogether with the number of Thymidines (T), and thereby of Uracils (U), in the polypyrimidine tract within the intron that facilitate spliceosomal assembly (Krchnakova et al., 2019). We analyzed splicing at seedlings stage and, taking into account the observation that *lncNATs-UGT73C6* expression vary during growth (Figure 8 C), it could also be possible that splicing of *lncNATs-UGT73C6* itself may be subject to developmental regulation. This could provide or limit sequence specific interactions with regulatory factors, with respect to changing physiological contexts or potential stress conditions similar to reported stress-specific lncRNAs in *A. thaliana* (Calixto et al., 2019).

Nonetheless, characteristic organ specific spatiotemporal expression of *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* further strengthened our assumptions that *lncNATs-UGT73C6* could play biologically significant role in *A. thaliana*. Due to the lack of information concerning the promoter region necessary for *lncNATs-UGT73C6* transcription initiation and to avoid the exclusion of potential upstream regulatory elements, our reporter *GUS* constructs included up-to 2.5 kb sequence upstream of reported transcription initiation sites of *lncNATs-UGT73C6*. Since weaker *lncNAT1* and stronger *lncNAT2* expression is restricted to roots and shoot tissues respectively (Figure 8 A and Supplementary Figure 1), further mapping of distinct or shared *lncNAT1* and *lncNAT2* promoter sequences may shed light not only over the minimal promoters but also to enable further investigations about regulatory elements and related transcription factors that specifies different expression patterns of *lncNATs-UGT73C6*. However, given the sensitivity and limitations of reporter GUS assays (Taylor, 1997), it could also be possible that both *lncNATs-UGT73C6* are expressed throughout plant organs to different degrees in

various tissues and it might be the relative ratio that is relevant for organ or cell type specific functions as shown for *Klhl14-AS* antisense lncRNA in mouse (Credendino et al., 2017). Although we substantiated our reporter GUS assay results with other approaches e.g. RT-PCR & qPCR, the observation that root specific *lncNAT1* is also expressed in shoots (data not shown) supports the above-mentioned assumption. Nevertheless, tissue specific $_{prom}$::GUS activity, cytoplasmic localization, longer RNA half-life ($t_{1/2}$) and developmentally regulated expression pattern of *lncNATs-UGT73C6* in *A. thaliana* laid down the foundations for their functional characterization (Figure 8 and Supplementary Figure 1).

## 6.2 Role *lncNATs-UGT73C6* in *A. thaliana* leaf development

Growth-stage based analysis (Boyes et al., 2001) of developmental phases showed that phenotypic parameters such as root length, fresh weight and seed yield are not affected due to alteration of *lncNATs-UGT73C6* levels (Supplementary Figure 13). Also, preliminary experiments indicated that the emergence of inflorescence, opening of first flower, and length of stems does not differ in overexpression or knockdown lines of *lncNATs-UGT73C6* compared to control plants. Nonetheless, constitutive overexpression of *lncNATs-UGT73C6* (either *lncNAT1* or *lncNAT2*) showed changes in leaf area both globally and locally at rosette and individual leaf level respectively (Figure 9, 10, 11 D and 13 A). As mentioned in the result section 5.1, *lncNAT2* sequence is included in *lncNAT1* transcripts and *lncNAT2* promoter is specifically active in leaf tissues, our priority was to elucidate role of *lncNAT2* in detail. Few remarks are critical to the observed effects of *lncNAT2* at leaf size viz: i) the magnitude of reproducible phenotypic effects is relatively less strong in contrast to the reported effects of the overexpression of the protein coding genes *UGT73C6* or *UGT73C5* (Poppenberger et al., 2005; Husar et al., 2011)*,* ii) the morphological changes in leaf area are rather accumulative and becomes more apparent during later developmental stages (Supplementary Figure 4), and iii) the contribution of each rosette leaf in the overall globally quantified rosette area is different. Although the identification of molecular processes that limits the strength of the effects for *lncNATs-UGT73C6* over final leaf size is beyond the scope of this thesis, review of literature highlights that degree of phenotypic strength for *lncNATs-UGT73C6* is comparable to the magnitude described for other plant NAT-lncRNAs. It is reported that *COOLAIR, asDOG1, MAS, asHSFB2a etc.* does not exhibit strong phenotypic effects. Expression of *COOLAIR* does not does not result in absolute inhibition of *FLC* expression (Swiezewski et al., 2009) implicating the simultaneous interplay of other regulatory factors in the regulation of flowering time control in response to the cold. *asDOG1* is reported to play role in the

germination as well as in drought stresses (Fedak et al., 2016; Yatusevich et al., 2017). It is shown that *asDOG1* expression is downregulated and *DOG1* is upregulated in T-DNA insertional mutant of *DOG1* allele *(dog1-5)*. 25% of *dog1-5* seeds are still able to propagate indicating complex regulation of germination process (Fedak et al., 2016). Also, ~50% downregulation of cold induced *MAS* by amiRNAs in *trans* shows that amplitude of early flowering phenotype is close to 20% (Zhao et al., 2018). Moreover complete downregulation of *HSFB2a* by ectopic overexpression of its *antisense* causes sterility only in ~45% of the ovules (Wunderlich et al., 2014). These examples highlight that developmental effects of studied lncRNAs are probably outcomes of very complex and intricate molecular processes which involve multiple regulatory factors. Therefore, it could be possible that *lncNATs-UGT73C6* might play fine tuning roles in the overall regulation of *A. thaliana* leaf size in conjunction with other possible molecular processes listed in leaf the development section 2.3.

Using a highly specific amiRNA approach (Carbonell et al., 2014) we could only partially downregulate *lncNATs-UGT73C6*. Yet, 50 to 60% downregulation of *lncNATs-UGT73C6* in *trans* was sufficient to induce developmental effects on rosette area (Figure 11 D). In agreement, low efficiency of amiRNA approach for the downregulation of lncRNAs can be observed in case of published results over *MAS*. Using the *mir159* backbone for the amiRNA construct, Zhao et.al. showed downregulation in the levels of *MAS*, which is nuclear localized and interacts with WDR5a protein, by a factor close to half (Zhao et al., 2018). It could be possible that the *amiRNA* mediated silencing in cytoplasm could further worsen if cytosolic *lncNATs-UGT73C6* mediate their action forming ribonucleoprotein complexes, thereby limiting accessibility to the target site in *lncNATs-UGT73C6* molecules. At the same time most of lncRNAs possess small ORFs and could be subjected to a RNA surveillance mechanism called nonsense mediated decay (NMD) due to presence of premature termination codons (Kurihara et al., 2009). However, lncRNAs will be subject to translation and NMD only when they do not form dsRNA structures and if the interacting proteins dissociate from lncRNA molecules (Wery et al., 2016). Furthermore, the observed higher stability of *lncNATs-UGT73C6* (Figure 8 D), thus, is an additional indication that *lncNATs-UGT73C6* might interact with proteins supporting the above-mentioned notion regarding reduced efficiency of amiRNA approach. Moreover, further inefficient downregulation of *lncNATs-UGT73C6* in *amiRNA2* and *amiRNA3* lines (Figure 11 B), compared to *amiRNA1* construct could also be due to reduced processing of amiRNA precursors (Figure 11 A). In addition, dynamic tertiary structure of *lncNATs-UGT73C6* could adversely affect cleavage activity of amiRNA guided AGO complex. Thus, a number of factors can influence competence of amiRNA approach. In

71

downregulation lines mature amiRNA are 21nt long and only the guide strand is loaded onto the AGO complex. Therefore the possibility of amiRNA mediated triggering of potential secondary siRNAs (Carbonell, 2019; de Felippes, 2019) that may induce undesired silencing of the sense *UGT73C6* gene or of other closely related *UGT73C* family members can also be ruled out. Further validation by small RNA sequencing could have been performed for these extreme possibilities. However, in our case the expression levels of *UGT73C6* and other family members remained unaffected due to *lncNATs-UGT73C6* downregulation (Figure 17). Nevertheless, to our knowledge it was the only optimal and cleanest means available to address the functionality *lncNATs-UGT73C6* because traditional RNAi approach (Bernards, 2006; Travella and Keller, 2009) using hairpins has its own limitations particularly for overlapping NAT-lncRNAs. Use of hairpins would cause downregulation of both sense and antisense gene simultaneously and therefore it will not be possible to distinguish potential downregulation effects of NAT-lncRNA from cognate protein coding gene. Although method is widely successful for studying protein coding genes, the success of RNAi based knockdown of lncRNAs, other than head to tail NAT-lncRNAs, is also known to be affected by multiple factors such as localization, low endogenous expression levels, folding and self-base pairing of lncRNA, accessibility of target site, etc. Moreover, RNAi can cause off-target effects and therefore the risk of affecting expression of neighboring genes is also associated (Charles Richard and Eichhorn, 2018).

In parallel, we also investigated the role of *lncNATs-UGT73C6* by means of overexpression. It has been shown for protein coding genes that the overexpression can result in promiscuous interactions, resources limitations, stoichiometric imbalance and pathway modulation thereby affecting interpretations of results (Moriya, 2015). Notwithstanding, overexpression strategy is a widely employed tool to analyze the functionality of lncRNAs (Mattick, 2009; Prelich, 2012; Kashi et al., 2016; Charles Richard and Eichhorn, 2018; Liu and Lim, 2018). Results obtained after alteration in levels of *lncNATs-UGT73C6* did not show a perfect correlation for the expression levels of *lncNATs-UGT73C6* with the phenotypic output, which was consistent in repeatedly performed experiments (Table 2 to 4), in overexpression and downregulation lines. While overexpression of *lncNATs-UGT73C6* in different transgenic lines may have saturating effects at lower threshold beyond which *lncNATs-UGT73C6* cannot affect developmental outcome, the different extents of downregulation in independent *amiRNA* lines did not result in corresponding decrease in leaf area (Figure 11 B and Table 5). Variation in quantified rosette area, despite a striker control of growth conditions, within the population of respective genotype could be attributed to different degree of overexpression or downregulation in the individual plants

in respective transgenic line. It is also important to consider that UGTs are result of recent gene duplications events during evolution and that described role of *lncNATs-UGT73C6*, in the multigene family context, might be rather complementary in combination with other regulatory pathways determining leaf shape and size. Therefore, phenotypic effects due to downregulation might be associated with imbalance in abundance instead of being dependent over the critical threshold expression levels of *lncNATs-UGT73C6*. Also, from evolutionary perspectives, it could be that *lncNATs-UGT73C6* only fine-tunes the mechanism(s) controlling leaf size in a manner similar to developmental roles of other lncRNAs in animals (Dahl et al., 2018) and plants (Datta and Paul, 2019). Nevertheless, based upon complementary results obtained from the overexpression and downregulation studies, in terms of changes in rosette area, we report, to our knowledge the first NAT-lncRNA that play developmental role in *trans* to modulate leaf size in *A. thaliana*.

At the cellular level *lncNAT2* overexpression showed localized effect only at the leaf bottom which also coincides with *lncNAT2$_{prom}$::GUS* activity pattern (Figure 23). The results indicate that *lncNAT2* play role during early developmental phases of leaf morphogenesis that are mostly accompanied by coordinated processes of cell division, elongation and expansion. *LncNAT2$_{prom}$::GUS* activity gradually decreases as leaf matures and confines at the basal part of leaf primordia (Figure 23). This pattern is similar to the mesophyll specific *an3-4/promAN3::GUS* activity shown for the first leaf in 6 days old seedlings (Kawade et al., 2013) suggesting that *lncNAT2* could have role in processes similar to those described for *AN3* (Tsukaya, 2013b). Based upon the rough approximation of total cell numbers in entire leaf, it appears that *lncNAT2* probably participate in coordinated regulation of cell proliferation and/or cell expansion. Nonetheless, detailed kinematic analysis for leaf growth and cellular parameters are further required for deeper characterization of *lncNAT2* function.

## 6.3  Potential mechanism(s) of *lncNATs-UGT73C6* action

### 6.3.1 Absence of local gene expression regulatory loop between *lncNATs-UGT73C6* and *UGT73C6*

*LncNATs-UGT73C6* are encoded from the complementary DNA strand of *UGT73C6*, a member of the *UGT73C* subfamily in the D group of plant UGTs. Our initial idea was based upon the hypothesis that *lncNATs-UGT73C6* can form an anticipated locus-specific feedback loop to regulate expression of *UGT73C6* and other family members via siRNA mediated gene silencing

(Borsani et al., 2005; Held et al., 2008). Due to absence of histone modification marks in *UGT73C6* promoter (Araport), we also did not consider above prospects of a *lncNATs-UGT73C6* driven regulatory mechanism to control expression of multiple and homologues genes similar to *AS1DHRS4* antisense lncRNA in mammals (Li et al., 2012). In fact our previous extensive attempts using transient expression assay in *N. benthamiana* concurred with above assumption regarding the mechanism of *lncNATs-UGT73C6* via siRNA pathway (de-Vries, 2014). However, alteration in *lncNATs-UGT73C6* did not show expected changes in levels of *UGT73C6* or other family members in stably transformed transgenic lines of *A. thaliana* (Figure 17). One of the main reasons could be that transient analysis in *N. benthamiana* does not accurately represent real *in vivo* scenario of *Arabidopsis* and that several developmental stage specific regulatory factors might be missing or be different in *N. benthamiana*. Surprisingly *UGT73C6* expression levels are slightly elevated upon overexpression of *lncNATs-UGT73C6* indicating a potential stabilizing effects of *lncNATs-UGT73C6* over relatively less stable *UGT73C6* mRNA.  Potential formation of dsRNAs between *lncNATs-UGT73C6* and *UGT73C6* can lead to unexpected elevation in abundance of *UGT73C6*. However, if why dsRNAs are formed and they are not cleaved by DICER machinery is topic of further investigations. Processing of dsRNA by DICER can be subject of regulation by other factors which can influence cleavage activity and efficiency including that of the structure of dsRNA itself (Vermeulen et al., 2005). Since downregulation of *lncNATs-UGT73C6* did not affect *UGT73C6* expression, we did not further investigate the physiological relevance of the slight increase in *UGT73C6* levels due to overexpression of *lncNATs-UGT73C6*.

Considering that *UGT73C5* and *UGT73C6* play role in BR homeostasis (Poppenberger et al., 2005; Husar et al., 2011), we also tested the possibility of whether *lncNATs-UGT73C6* can form a BR driven feedback loop with *UGT73C6.* Since *lncNATs-UGT73C6* are only slightly induced during particular time points upon BR treatment unlike a uniform alterations in BR specific control genes *BAS1* (Figure 18 A), *BR6ox2* and *DWF4* (Supplementary Figure 10 B), we are suspicious that induction of *lncNATs-UGT73C6* for a very short duration has any particular physiological relevance. Terminal effects of altered BR levels *in planta* are best studied by quantifying changes in the length of etiolated hypocotyl (Chung and Choe, 2013; Espinosa-Ruiz et al., 2015). No obvious alterations in the hypocotyl length or in the transcript levels of BR biosynthetic (*ROT3*) and signaling (*EXP8*) genes upon overexpression or downregulation of *lncNATs-UGT73C6* further strengthen our above interpretation (Figure 19 A to C). With respect to developing leaves, prom::GUS activity of *UGT73C6* or *UGT73C5* does not show overlapping expression pattern with the reporter activity for *lncNAT2* and the promoters of *lncNATs-UGT73C6* are able to respond

independent of their genomic locus (Figure 8 A and Supplementary Figure 1). Unless *lncNATs-UGT73C6* are mobile RNA molecules similar to other examples of lncRNAs in *Arabidopsis* (Zhang et al., 2019), collectively these data rules out the possibility that identified role of *lncNATs-UGT73C6* in leaf development is mediated via the gene expression regulation of *UGT73C6* or *UGT73C5*. Although *lncNAT1* reporter activity in roots overlaps with strongly expressed *UGT73C5* and poorly expressed *UGT73C6*, therefore, it would be interesting to further investigate the root specific potential gene regulatory mechanisms between *lncNAT1*, *UGT73C6* and *UGT73C5*. In contrast, the overexpression of *lncNAT1* produces phenotype like *lncNAT2* in leaves (Figure 9 and 10), our focus was over the functional characterization of *lncNAT2*. Additionally, why *UGT73C6* remains poorly expressed during most of development compared to *lncNAT2* and *UGT73C5* (Figure 8 C) remains to be understood. Interestingly, the levels of upregulation for *UGT73C6, UGT73C5* and *lncNATs-UGT73C6,* after DON treatment, are comparable which suggest that these genes may have role in responses related to pathogen stress (Figure 20 B, C & D). Contrary to DON assay results, pattern of oscillations in expression levels of *UGT73C5* and *lncNATs-UGT73C6* (preassembly *lncNAT2*) suggest that a possibility of connected developmental role cannot be excluded (Figure 8 C). Equally, the overexpression of *UGT73C5* has no affect over *lncNATs-UGT73C6* (Figure 15 F) and alterations in expression levels of *lncNATs-UGT73C6* does not change expression levels of *UGT73C5* (Figure 17). We perceive that role of *lncNATs-UGT73C6* in leaf development is not mediated via regulation of *UGT73C5*. However, given that *UGT73C5* is a recent duplicate in *A. thaliana* genome (Ross et al., 2001), plays a role in BR homeostasis (Poppenberger et al., 2005), it might still be in the process of evolution and functionalization (Wang et al., 2016). Therefore, the observed coordinated oscillations in the expression level of *lncNATs-UGT73C6* with *UGT73C5* during growth phases (Figure 8 C) may indirectly be needed to counter-balance BR deficiency supposedly imposed by promiscuously active UGT73C5. This is a mere speculation and postulation needs to be thoroughly tested from several standpoints. Besides, out of the genomic context, promoters of *lncNAT2* are unresponsive while of *lncNAT1* is slightly induced by DON treatment. Data indicate that transcriptional activity over *lncNAT2* promoters may not be a primary source of DON induced upregulation in the expression levels of *lncNATs-UGT73C6*. On the other hand, *UGT73C6* overexpression as well, like DON treatment, induces accumulation of unspliced *lncNAT1* potentially due to RDR activity. Considering that *UGT73C6* plays role in stress (Figure 20 B), *lncNAT1* might be physiologically relevant to clean up *UGT73C6* mRNA after the stress as suggested in cases of NAT-lncRNAs (Matsui et al., 2017) or to probably reduce the affects over leaf growth when *UGT73C6* is in excess

(Figure 15 A to E). How transiently accumulated unspliced variant of *lncNAT1* circumvent or coordinates with the developmental role *lncNAT2* when *UGT73C6* is induced, remained to be investigated. In contrast, apart from the small-scale differences (28 base pairs over antisense strand of *UGT73C5* corresponding to assumed *lncNAT2* promoter sequence) in the 3' ends of *UGT73C5* and *UGT73C6* (Supplementary Figure 14), the association of *lncNATs-UGT73C6* with *UGT73C6* over opposite DNA strand might have consequential effects. In comparison to *UGT73C5*, the act of *lncNATs-UGT73C6* transcription itself might be a causality for maintaining poor expression of *UGT73C6* during developmental phases (Figure 8 C) implying a regulatory module in *cis* (Pelechano and Steinmetz, 2013).

## 6.3.2 Effects of *lncNATs-UGT73C6* knockout in *cis*

Unlike downregulation effects of *lncNAT2* in *trans*, CRISPR/Cas9 mediated complete or partial knockout of *lncNATs-UGT73C6* in *Cr_lncnats #13*.3 and *Cr_lncnats #18*.2 lines did not result in expected phenotypic changes over rosette area (Figure 21 D). On the other hand, *amiRNA* lines shows reduced rosette area in plants older than 25 days (Table 5). The *cis* knockout of *lncNATs-UGT73C6* rather results in early developmental defects in seedlings marked by the presence of asymmetric cotyledons and trilateral symmetry of cotyledons in a small part of the population, phenotypes that are similar to the reported knockout lines of the regulatory component of *mir396-GRFs* module (Kanei et al., 2012) (Figure 21 E). The frequency of early developmental defects was higher in *Cr_lncnats #18*.2 lines as shown (Figure 21 E) from ~150 seedlings grown over ½ MS agar plates. Moreover, similar defects were also observed for *Cr_lncnats #1*3.3 when the number of seeds was raised up to ~1000 (Mr. M. Heidecker, personal communication). Sequence analysis of *Cr_lncnats #13*.3 and *Cr_lncnats #18.2* lines further confirmed that growth defects in cotyledons are not due to mutations in the *AN3* locus (Mr. M. Heidecker, personal communication). The possibility off-target mutations of other UGTs is also highly unlikely because the gRNA used in CRISPR/Cas9 editing of *lncNATs-UGT73C6* are quite specific. Interestingly, as mentioned before, *cis* knockout does not produce the effects observed in *trans* and vice versa. Since, it is known that final organ leaf size is outcome of complex processes that take place over a period of time during early leaf growth (Gonzalez et al., 2012), the continuous absence of *lncNATs-UGT73C6* in knockout lines might trigger ameliorative compensatory pathways soon after emergence of leaf primordia (Tsukaya, 2003; Horiguchi et al., 2006). Nonetheless, few remarks are crucial in this regard. Mechanisms of lncRNAs are far diverse than originally thought and the possibility of a multifaceted regulatory mechanism for *lncNATs-UGT73C6* that

simultaneously includes both *cis* and *trans* aspects cannot be excluded. Importantly, not all, but in some cases it has been shown that lncRNA locus functions in *cis* while its lncRNA transcript can play different role in *trans* (Paralkar et al., 2016). We generated complete and partial knockout lines for *lncNATs-UGT73C6*. In one of our CRISPR/Cas9 knockout line i.e. *Cr_lncnats #18.2,* 603 bp long deletion was introduced at 3' end (+1229 to +1795 and an insertion-deletion of 56 bp replacing 93 bp) allowing the preservation of 5' upstream DNA sequence close to the *lncNATs-UGT73C6* promoters. In *Cr_lncnats #13*.3 line, complete deletion (+45 to +1795) of *lncNATs-UGT73C6* is introduced including that of 5' region in *UGT73C6*. Unlike *Cr_lncnats #13*.3, in *Cr_lncnats #18.2* truncated *lncNATs-UGT73C6* are transcribed similar to *Col-0* or *amiRNA* lines. Despite the fact, deletion of the genomic sequence is introduced in the region that is required for transcriptional initiation of cognate protein coding gene leading to the imbalance in the ratio of *lncNATs-UGT73C6* and *UGT73C6*. Although unavoidable, this might be particularly important if *lncNATs-UGT73C6* play a role in *cis,* that is dependent over the integrity of locus, for the potential regulation of *UGT73C6* and/or *UGT73C5* sister pair. Indeed *cis* regulatory elements are suggested to be crucial in the evolution of regulatory networks for duplicated genes (Arsovski et al., 2015). This might be a reason that can be attributed to variation in results obtained after the downregulation of *lncNATs-UGT73C6* by CRISPR/Cas9 and amiRNA approach. In contrast, downregulation in *trans* by *amiRNAs* takes place post-transcriptionally only after the act of transcription, splicing and follow up localization of *lncNATs-UGT73C6* in cytoplasm. Importantly, the transcriptional activity of *lncNATs-UGT73C6* and *UGT73C6* is not affected in *amiRNA* lines. Also, whether integrity of the locus, transcription, splicing and downstream regulatory processes are together linked for the role of *lncNATs-UGT73C6* remains to be investigated. Undoubtedly, in the purview of complex regulatory mechanisms of lncRNAs, it is important to jointly consider various aspects of *cis* and *trans* while investigating the transcript-based and/or transcription-based effects of lncRNAs. Given the wide-ranging diversity in molecular mechanisms of lncRNAs, all-inclusive consideration of the different data sets is extremely crucial in the overall interpretation of results obtained from different approaches. Experimental manipulations of gene expression in *cis* and *trans* may not necessarily work in similar fashion (Signor and Nuzhdin, 2018). Therefore, further studies are needed to address the differences in the results obtained after the downregulation of *lncNATs-UGT73C6* in *cis* and *trans*.

### 6.3.3 Potential target mimicry of *miR396* and evolutionary significance of *lncNAT2*

Functional studies of *lncNAT2* clearly indicate that it is a bona fide lncNAT-lncRNA and its effect over leaf area is independent of *UGT73C6* transcript levels. As mentioned in section 2.3, several gene regulatory pathways play role in the leaf development. Among others, *mir396-GRF* regulatory module (Kim et al., 2003; Kim and Lee, 2006; Lee et al., 2009; Rodriguez et al., 2010) is critical for final leaf size determination. *mir396* targets *GRFs* that positively affect leaf size. Cytosolic *lncNAT2* shows sequence complementarity with *miR396*. Interestingly, alignment of RNA sequences shows formation of bulge loop between *lncNAT2:miR396* hybrid at the nucleotide position essential for the cleavage activity of *miR396*-AGO complex (Figure 23). Also, the promoter activity localization of *GRFs* such as *GRF2* (Debernardi et al., 2012) is similar to that of *lncNAT2* (Figure 23). In addition, the effect of *lncNAT2* overexpression at the cellular level is quite specific and stronger in mesophyll cells only at the bottom of leaf in accordance with reporter activity. Not performed yet, however it would be interesting to analyze if there are also changes at microscopic level due to downregulation of *lncNATs-UGT73C6* in *trans*. Additionally, *cis* knockout of *lncNATs-UGT73C6* results in early developmental effects similar to the *an3* mutant (Kanei et al., 2012). Based upon these observations, we hypothesize that *lncNAT2* acts as a target mimic, that could modulate *GRFs* levels by sequestering *miR396* to positively regulate leaf development in *A. thaliana* (Figure 23).

In support of the above hypothesis, we quantified *GRFs* levels in *lncNAT2* downregulation lines. Contrarily to what expected, *GRFs* levels were not drastically affected and only a marginal change in *GRFs* abundance could be observed. Only *GRF2* was found to be slightly affected (Figure 22). This could be due to due to use of whole seedlings used for total RNA extractions. Though its challenging to physically dissociate and quantify leaf region where *lncNAT2* functions, a local mapping of changes in transcripts levels of *GRFs* could be carried out in future by finely dissecting leaves into bottom and tip regions. Additionally, RNA seq analysis can be carried out from these samples to assess the global transcriptional changes in various transgenic lines harboring overexpression or knockout of *lncNAT2*. *In vitro* cleavage assays using *Nicotiana tabacum* BY-2 cell lysates, the BYL system, (Gursinsky et al., 2009; Gago-Zachert et al., 2019) by employing *lncNAT2*, *miR396* and target *GRF* transcripts could also help to elucidate postulated effects of *lncNAT2* over *miR396* and *GRFs*. Nevertheless, an imperfect target mimicry of *miR396* by *lncNAT2* could be biologically relevant. *UGT73C6* and *UGT73C5* are recently evolved sister

pairs and share high level of sequence similarity. Proteins encoded by both genes are reported to be promiscuous for their enzymatic properties and their overexpression affects BR homeostasis (Poppenberger et al., 2003; Poppenberger et al., 2005; Husar et al., 2011) (Figure 15). Furthermore, promiscuous catalytic properties of UGT73C6 and UGT73C5 is also supported by the fact that both genes are induced by DON (Husar et al., 2011), however only UGT73C5 was shown to inactivate DON (Schweiger et al., 2010). As stated earlier in section 6.3.1, functionalization of these homologues may still not be achieved (Ross et al., 2001). Conversely, the regulation of BR homeostasis is rather an elaborated and tightly controlled process involving well conserved regulatory cascades across plant species both for biosynthesis and signaling genes (Zhu et al., 2013). An insufficient target mimicry of *miR396* by the non-conserved *lncNAT2* may be relevant not only for counterbalancing the BR deficiency caused by promiscuous *UGT73C6* and/or *UGT73C5* but also to positively facilitate evolution. *LncNAT2* is specific to *A. thaliana* like many other species-specific lncRNAs (Paralkar et al., 2014). Surprisingly, antisense strand of *UGT73C5* does not encode for a NAT-lncRNA. Comparison of *lncNAT2* promoter sequence, starting from 958 bp until 1440 bp downstream of *UGT73C6* transcription start site with respective sequence at 3' region of UGT73C5 shows 28 bp sequence dissimilarity and the differences are scattered (Supplementary Figure 14). It could be possible that association of *lncNATs-UGT73C6* only with one of the sister pair might have consequential effects in *cis* resulting in poor basal expression of *UGT73C6*, while allowing normal expression of *UGT73C5* during the developmental. Under these circumstances and during the course of evolution, *lncNAT2* might have evolved in species-specific manner to compensate the undesired effects of *UGT73C5* over BR levels via imperfect target mimicry of *miR396* to increase the levels of *GRFs* in order to prevent inhibitory effects over leaf growth (Figure 23). In this manner, *lncNATs-UGT73C6* might be important for the overall regulation of dosage response of *UGT73C6* and *UGT73C5.* Also, such NATs-lncRNAs mediated regulatory cascades in the context of multigene families might be a general phenomenon across organisms for the functional specialization and co-evolution of the sister pairs simultaneously.
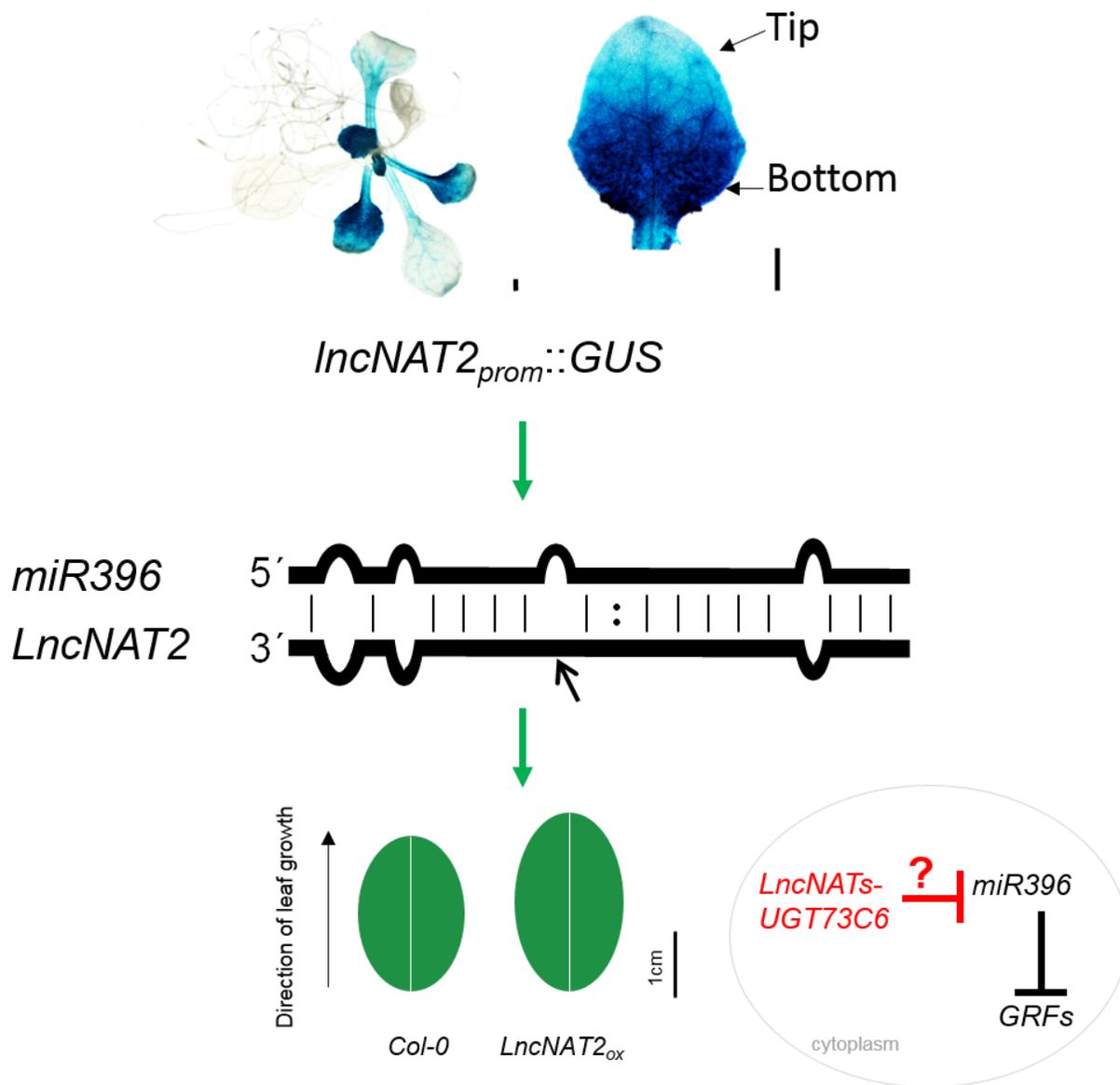
Figure 23: **Potential mode of action of** *lncNATs-UGT73C6*. *LncNAT2* is expressed in *A. thaliana* leaves at bottom part during early leaf development (top). In the cytoplasm, *lncNAT2* could potentially bind to *miR396* (middle) leading to elevation in levels of *GRFs* to positively regulate leaf growth (bottom).

# 7 References

**Abel, S., Savchenko, T., and Levy, M.** (2005). Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and Oryza sativa. BMC Evol Biol **5,** 72.

**Achard, P., and Genschik, P.** (2009). Releasing the brakes of plant growth: how GAs shutdown DELLA proteins. J Exp Bot **60,** 1085-1092.

**Achard, P., Gusti, A., Cheminant, S., Alioua, M., Dhondt, S., Coppens, F., Beemster, G.T., and Genschik, P.** (2009). Gibberellin signaling controls cell proliferation rate in Arabidopsis. Curr Biol **19,** 1188-1193.

**Anastasiou, E., Kenz, S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C., and Lenhard, M.** (2007). Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Dev Cell **13,** 843-856.

**Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Muhlenbock, P., Skirycz, A., Gonzalez, N., Beemster, G.T., and Inze, D.** (2012). Exit from proliferation during leaf development in Arabidopsis thaliana: a not-so-gradual process. Dev Cell **22,** 64-78.

**Aparicio-Prat, E., Arnan, C., Sala, I., Bosch, N., Guigo, R., and Johnson, R.** (2015). DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC genomics **16,** 846.

**Ariel, F., Romero-Barrios, N., Jegu, T., Benhamed, M., and Crespi, M.** (2015). Battles and hijacks: noncoding transcription in plants. Trends in plant science **20,** 362-371.

**Arsovski, A.A., Pradinuk, J., Guo, X.Q., Wang, S., and Adams, K.L.** (2015). Evolution of Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of Arabidopsis. Plant physiology **169,** 2982-2991.

**Autran, D., Jonak, C., Belcram, K., Beemster, G.T., Kronenberger, J., Grandjean, O., Inze, D., and Traas, J.** (2002). Cell numbers and leaf development in Arabidopsis: a functional analysis of the STRUWWELPETER gene. EMBO J **21,** 6036-6049.

**Azpiroz, R., Wu, Y., LoCascio, J.C., and Feldmann, K.A.** (1998). An Arabidopsis brassinosteroid-dependent mutant is blocked in cell elongation. The Plant cell **10,** 219-230.

**Barrell, B.G., Air, G.M., and Hutchison, C.A., 3rd.** (1976). Overlapping genes in bacteriophage phiX174. Nature **264,** 34-41.

**Ben Amor, B., Wirth, S., Merchan, F., Laporte, P., d'Aubenton-Carafa, Y., Hirsch, J., Maizel, A., Mallory, A., Lucas, A., Deragon, J.M., Vaucheret, H., Thermes, C., and Crespi, M.** (2009). Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome research **19,** 57-69.

**Bernards, R.** (2006). [The Nobel Prize in Physiology or Medicine for 2006 for the discovery of RNA interference]. Ned Tijdschr Geneeskd **150,** 2849-2853.

**Björkman, O.** (1981). Responses to Different Quantum Flux Densities. In Physiological Plant Ecology I: Responses to the Physical Environment, O.L. Lange, P.S. Nobel, C.B. Osmond, and H. Ziegler, eds (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 57-107.

**Borges, F., and Martienssen, R.A.** (2015). The expanding world of small RNAs in plants. Nature reviews. Molecular cell biology **16,** 727-741.

**Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K.** (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell **123,** 1279-1291.

**Bowles, D., Isayenkova, J., Lim, E.K., and Poppenberger, B.** (2005). Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol **8,** 254-263.

# References

**Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., and Gorlach, J.** (2001). Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. The Plant cell **13,** 1499-1510.

**Brosius, J.** (2005). Waste not, want not--transcript excess in multicellular eukaryotes. Trends in genetics : TIG **21,** 287-288.

**Bryant, D.A., and Frigaard, N.U.** (2006). Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol **14,** 488-496.

**Burstenbinder, K., Moller, B., Plotner, R., Stamm, G., Hause, G., Mitra, D., and Abel, S.** (2017). The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus. Plant physiology **173,** 1692-1708.

**Calixto, C.P.G., Tzioutziou, N.A., James, A.B., Hornyik, C., Guo, W., Zhang, R., Nimmo, H.G., and Brown, J.W.S.** (2019). Cold-Dependent Expression and Alternative Splicing of Arabidopsis Long Non-coding RNAs. Front Plant Sci **10,** 235.

**Candelli, T., Challal, D., Briand, J.B., Boulay, J., Porrua, O., Colin, J., and Libri, D.** (2018). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. EMBO J **37**.

**Carbonell, A.** (2019). Secondary Small Interfering RNA-Based Silencing Tools in Plants: An Update. Front Plant Sci **10,** 687.

**Carbonell, A., Takeda, A., Fahlgren, N., Johnson, S.C., Cuperus, J.T., and Carrington, J.C.** (2014). New generation of artificial MicroRNA and synthetic trans-acting small interfering RNA vectors for efficient gene silencing in Arabidopsis. Plant physiology **165,** 15-29.

**Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T.,

**Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., Consortium, F., Group, R.G.E.R., and Genome Science, G.** (2005). The transcriptional landscape of the mammalian genome. Science **309,** 1559-1563.

**Casal, J.J.** (2012). Shade avoidance. Arabidopsis Book **10,** e0157.

**Cech, T.R.** (2012). The RNA worlds in context. Cold Spring Harbor perspectives in biology **4,** a006742.

**Cerase, A., Pintacuda, G., Tattermusch, A., and Avner, P.** (2015). Xist localization and function: new insights from multiple levels. Genome biology **16,** 166.

**Charles Richard, J.L., and Eichhorn, P.J.A.** (2018). Platforms for Investigating LncRNA Functions. SLAS Technol **23,** 493-506.

**Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J.R., and Belostotsky, D.A.** (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell **131,** 1340-1353.

**Chen, L.L.** (2016). Linking Long Noncoding RNA Localization and Function. Trends Biochem Sci **41,** 761-772.

**Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D.** (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant journal : for cell and molecular biology **89,** 789-804.

**Choe, S., Fujioka, S., Noguchi, T., Takatsuto, S., Yoshida, S., and Feldmann, K.A.** (2001). Overexpression of DWARF4 in the brassinosteroid biosynthetic pathway results in increased vegetative growth and seed yield in Arabidopsis. The Plant journal : for cell and molecular biology **26,** 573-582.

**Chung, Y., and Choe, S.** (2013). The Regulation of Brassinosteroid Biosynthesis in Arabidopsis. Critical Reviews in Plant Sciences **32,** 396-410.

**Clouse, S.D., Langford, M., and McMorris, T.C.** (1996). A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development. Plant physiology **111,** 671-678.

**Collins, T.J.** (2007). ImageJ for microscopy. BioTechniques **43,** 25-30.

**Cosgrove, D.J.** (2005). Growth of the plant cell wall. Nature reviews. Molecular cell biology **6,** 850-861.

**Credendino, S.C., Lewin, N., de Oliveira, M., Basu, S., D'Andrea, B., Amendola, E., Di Guida, L., Nardone, A., Sanges, R., De Felice, M., and De Vita, G.** (2017). Tissue- and Cell Type-Specific Expression of the Long Noncoding RNA Klhl14-AS in Mouse. Int J Genomics **2017,** 9769171.

**Crick, F.H.** (1958). On protein synthesis. Symposia of the Society for Experimental Biology **12,** 138-163.

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

**Cui, J., Luan, Y.S., Jiang, N., Bao, H., and Meng, J.** (2017). Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lncRNA16397 conferring resistance to Phytophthora infestans by co-expressing glutaredoxin. Plant Journal **89,** 577-589.

**Czech, B., Munafo, M., Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E., and Hannon, G.J.** (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annual review of genetics **52,** 131-157.

**da Rocha, S.T., and Heard, E.** (2017). Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. Nature structural & molecular biology **24,** 197-204.

**Dahl, M., Kristensen, L.S., and Gronbaek, K.** (2018). Long Non-Coding RNAs Guide the Fine-Tuning of Gene Regulation in B-Cell Development and Malignancy. Int J Mol Sci **19**.

**Datta, R., and Paul, S.** (2019). Long non-coding RNAs: Fine-tuning the developmental responses in plants. J Biosci **44**.

**Daviere, J.M., and Achard, P.** (2013). Gibberellin signaling in plants. Development **140,** 1147-1151.

**de-Vries, T.** (2014). Role of long non-coding RNAs in gene expression regulation of the *UGT73C* subfamily from *Arabidopsis thaliana*, Bachelor thesis 2014, Regulatory RNAs lab,  IPB, Halle (Saale).

**de-Vries, T.** (2016). Analysis of long non-coding RNAs and their interacting proteins. Master Thesis, Faculty of Natural Science I (Biological Science), Institute of Biology, Martin Luther University of Halle-Wittenberg

Regulatory RNAs lab, Department of Molecular Signal Processing, Leibniz-Institute of Plant Biochemistry (Halle), Germany**,** https://www.ipb-halle.de/en/research/molecular-signal-processing/research-groups/regulatory-rnas/.

**de Felippes, F.F.** (2019). Gene Regulation Mediated by microRNA-Triggered Secondary Small RNAs in Plants. Plants (Basel) **8**.

**Debernardi, J.M., Rodriguez, R.E., Mecchia, M.A., and Palatnik, J.F.** (2012). Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. PLoS genetics **8,** e1002419.

**Debernardi, J.M., Mecchia, M.A., Vercruyssen, L., Smaczniak, C., Kaufmann, K., Inze, D., Rodriguez, R.E., and Palatnik, J.F.** (2014). Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. The Plant journal : for cell and molecular biology **79,** 413-426.

**Deforges, J., Reis, R.S., Jacquet, P., Vuarambon, D.J., and Poirier, Y.** (2019a). Prediction of regulatory long intergenic non-coding RNAs acting in trans through base-pairing interactions. BMC genomics **20**.

**Deforges, J., Reis, R.S., Jacquet, P., Sheppard, S., Gadekar, V.P., Hart-Smith, G., Tanzer, A., Hofacker, I.L., Iseli, C., Xenarios, I., and Poirier, Y.** (2019b). Control of Cognate Sense mRNA Translation by cis-Natural Antisense RNAs. Plant physiology **180,** 305-322.

**Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J., and Guigo, R.** (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome research **22,** 1775-1789.

**Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M.B., Guo, Y., and Lu, Z.J.** (2014). Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. The Plant journal : for cell and molecular biology **80,** 848-861.

**Donnelly, P.M., Bonetta, D., Tsukaya, H., Dengler, R.E., and Dengler, N.G.** (1999). Cell cycling and cell enlargement in developing leaves of Arabidopsis. Dev Biol **215,** 407-419.

**Du, F., Guan, C., and Jiao, Y.** (2018). Molecular Mechanisms of Leaf Morphogenesis. Molecular plant **11,** 1117-1134.

**Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L.** (2010). MYB transcription factors in Arabidopsis. Trends in plant science **15,** 573-581.

**Dumais, J.** (2007). Can mechanics control pattern formation in plants? Current Opinion in Plant Biology **10,** 58-62.

**Easlon, H.M., and Bloom, A.J.** (2014). Easy Leaf Area: Automated digital image analysis for rapid and accurate measurement of leaf area. Appl Plant Sci **2**.

**Eddy, S.R.** (2012). The C-value paradox, junk DNA and ENCODE. Curr Biol **22,** R898-R899.

**Espinosa-Ruiz, A., Martínez, C., and Prat, S.** (2015). Protocol to Treat Seedlings with Brassinazole and Measure Hypocotyl Length in Arabidopsis thaliana. Bio-protocol **5,** e1568.

**Fan, J., Xing, Y., Wen, X., Jia, R., Ni, H., He, J., Ding, X., Pan, H., Qian, G., Ge, S., Hoffman, A.R., Zhang, H., and Fan, X.** (2015). Long non-coding RNA ROR decoys gene-specific histone methylation to promote tumorigenesis. Genome biology **16,** 139-139.

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

**Feng, G., Qin, Z., Yan, J., Zhang, X., and Hu, Y.** (2011). Arabidopsis ORGAN SIZE RELATED1 regulates organ growth and final organ size in orchestration with ARGOS and ARL. New Phytol **191,** 635-646.

**Foster, A.S.** (1936). Leaf Differentiation in Angiosperms. Botanical Review **Vol. 2**.

**Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J.** (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nature genetics **39,** 1033-1037.

**Fritz, M.A.** (2015). Analysis of the role of the natural antisense long non-coding RNAs complementary to UGT73C6 in gene expression regulation of the UGT73C subfamily. In Leibniz-Institut für Pflanzenbiochemie (Halle (Saale), Germany: Martin-Luther-Universität Halle-Wittenberg).

**Fujikura, U., Horiguchi, G., Ponce, M.R., Micol, J.L., and Tsukaya, H.** (2009). Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of Arabidopsis thaliana. The Plant journal : for cell and molecular biology **59,** 499-508.

**Gago-Zachert, S., Schuck, J., Weinholdt, C., Knoblich, M., Pantaleo, V., Grosse, I., Gursinsky, T., and Behrens, S.E.** (2019). Highly efficacious antiviral protection of plants by small interfering RNAs identified in vitro. Nucleic acids research **47,** 9343-9357.

**Gázquez, A., and Beemster, G.T.S.** (2017). What determines organ size differences between species? A meta-analysis of the cellular basis. New Phytol **215,** 299-308.

**Goethe, J.W.v.** (1790). Versuch die Metamorphose der Pflanzen zu erklären. CW Ettinger, Gotha, Germany **1st edition**.

**Golisz, A., Sikorski, P.J., Kruszka, K., and Kufel, J.** (2013). Arabidopsis thaliana LSM proteins function in mRNA splicing and degradation. Nucleic acids research **41,** 6232-6249.

**Gonzalez, N., Vanhaeren, H., and Inze, D.** (2012). Leaf size control: complex coordination of cell division and expansion. Trends in plant science **17,** 332-340.

**Gonzalez, N., De Bodt, S., Sulpice, R., Jikumaru, Y., Chae, E., Dhondt, S., Van Daele, T., De Milde, L., Weigel, D., Kamiya, Y., Stitt, M., Beemster, G.T., and Inze, D.** (2010). Increased leaf size: different means to an end. Plant physiology **153,** 1261-1279.

**Gotoh, E., Suetsugu, N., Higa, T., Matsushita, T., Tsukaya, H., and Wada, M.** (2018). Palisade cell shape affects the light-induced chloroplast movements and leaf photosynthesis. Sci Rep **8,** 1472.

**Guo, M., Thomas, J., Collins, G., and Timmermans, M.C.** (2008). Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of Arabidopsis. The Plant cell **20,** 48-58.

**Gursinsky, T., Schulz, B., and Behrens, S.E.** (2009). Replication of Tomato bushy stunt virus RNA in a plant in vitro system. Virology **390,** 250-260.

**Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L., and Lander, E.S.** (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458,** 223-227.

**Hamazaki, N., Nakashima, K., Hayashi, K., and Imamura, T.** (2017). Detection of Bidirectional Promoter-Derived lncRNAs from Small-Scale Samples Using Pre-Amplification-Free Directional RNA-seq Method. Methods in molecular biology **1605,** 83-103.

**Han, X., Luo, S., Peng, G., Lu, J.Y., Cui, G., Liu, L., Yan, P., Yin, Y., Liu, W., Wang, R., Zhang, J., Ai, S., Chang, Z., Na, J., He, A., Jing, N., and Shen, X.** (2018). Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. Journal of molecular cell biology **10,** 175-178.

**Harberd, N.P.** (2003). Botany. Relieving DELLA restraint. Science **299,** 1853-1854.

**Haroth, S., Feussner, K., Kelly, A.A., Zienkiewicz, K., Shaikhqasem, A., Herrfurth, C., and Feussner, I.** (2019). The glycosyltransferase UGT76E1 significantly contributes to 12-O-glucopyranosyl-jasmonic acid formation in wounded Arabidopsis thaliana leaves. The Journal of biological chemistry **294,** 9858-9872.

**Held, M.A., Penning, B., Brandt, A.S., Kessans, S.A., Yong, W., Scofield, S.R., and Carpita, N.C.** (2008). Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. Proc Natl Acad Sci U S A **105,** 20534-20539.

**Henriques, R., Wang, H., Liu, J., Boix, M., Huang, L.F., and Chua, N.H.** (2017). The antiphasic regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. New Phytol.

**Henz, S.R., Cumbie, J.S., Kasschau, K.D., Lohmann, J.U., Carrington, J.C., Weigel, D., and Schmid, M.** (2007). Distinct expression patterns of natural antisense transcripts in Arabidopsis. Plant physiology **144,** 1247-1255.

**Heo, J.B., and Sung, S.** (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science **331,** 76-79.

**Hibara, K., Karim, M.R., Takada, S., Taoka, K., Furutani, M., Aida, M., and Tasaka, M.** (2006). Arabidopsis CUP-SHAPED COTYLEDON3 regulates postembryonic shoot meristem and organ boundary formation. The Plant cell **18,** 2946-2957.

**Higgs, P.G., and Lehman, N.** (2015). The RNA World: molecular cooperation at the origins of life. Nature reviews. Genetics **16,** 7-17.

**Hisanaga, T., Kawade, K., and Tsukaya, H.** (2015). Compensation: a key to clarifying the organ-level regulation of lateral organ size in plants. J Exp Bot **66,** 1055-1063.

**Hombach, S., and Kretz, M.** (2016). Non-coding RNAs: Classification, Biology and Functioning. Advances in experimental medicine and biology **937,** 3-17.

**Horiguchi, G., Kim, G.T., and Tsukaya, H.** (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. The Plant journal : for cell and molecular biology **43,** 68-78.

**Horiguchi, G., Ferjani, A., Fujikura, U., and Tsukaya, H.** (2006). Coordination of cell proliferation and cell expansion in the control of leaf size in Arabidopsis thaliana. J Plant Res **119,** 37-42.

**Hou, B., Lim, E.K., Higgins, G.S., and Bowles, D.J.** (2004). N-glucosylation of cytokinins by glycosyltransferases of Arabidopsis thaliana. The Journal of biological chemistry **279,** 47822-47832.

**Hu, Y., Xie, Q., and Chua, N.H.** (2003). The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size. The Plant cell **15,** 1951-1961.

**Huang, S., Raman, A.S., Ream, J.E., Fujiwara, H., Cerny, R.E., and Brown, S.M.** (1998). Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in Arabidopsis. Plant physiology **118,** 773-781.

**Husar, S., Berthiller, F., Fujioka, S., Rozhon, W., Khan, M., Kalaivanan, F., Elias, L., Higgins, G.S., Li, Y., Schuhmacher, R., Krska, R., Seto, H., Vaistij, F.E., Bowles, D., and Poppenberger, B.** (2011). Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in Arabidopsis thaliana. BMC plant biology **11,** 51.

**Ichihashi, Y., Kawade, K., Usami, T., Horiguchi, G., Takahashi, T., and Tsukaya, H.** (2011). Key proliferative activity in the junction between the leaf blade and leaf petiole of Arabidopsis. Plant physiology **157,** 1151-1162.

**Ietswaart, R., Wu, Z., and Dean, C.** (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. Trends in Genetics **28,** 445-453.

**Ikeuchi, M., Yamaguchi, T., Kazama, T., Ito, T., Horiguchi, G., and Tsukaya, H.** (2011). ROTUNDIFOLIA4 regulates cell proliferation along the body axis in Arabidopsis shoot. Plant & cell physiology **52,** 59-69.

**Ingenhousz, J.** (1779). Experiments upon Vegetables, Discovering Their great Power of purifying the Common Air in the Sun-shine, and of Injuring it in the Shade and at Night. To Which is Joined, A new Method of examining the accurate Degree of Salubrity of the Atmosphere, London, 1779. . (McGraw Hill).

**Ito, T., Kim, G.T., and Shinozaki, K.** (2000). Disruption of an Arabidopsis cytoplasmic ribosomal protein S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development. The Plant journal : for cell and molecular biology **22,** 257-264.

**Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and 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. The Plant cell **25,** 4166-4182.

**Jain, A.K., Xi, Y., McCarthy, R., Allton, K., Akdemir, K.C., Patel, L.R., Aronow, B., Lin, C., Li, W., Yang, L., and Barton, M.C.** (2016). LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. Molecular cell **64,** 967-981.

**Jensen, T.H., Jacquier, A., and Libri, D.** (2013). Dealing with pervasive transcription. Molecular cell **52,** 473-484.

**Johnson, M.P.** (2016). Photosynthesis. Essays Biochem **60,** 255-273.

**Jones-Rhoades, M.W., and Bartel, D.P.** (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Molecular cell **14,** 787-799.

**Jones, P., Messner, B., Nakajima, J., Schaffner, A.R., and Saito, K.** (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. The Journal of biological chemistry **278,** 43910-43918.

**Kalve, S., De Vos, D., and Beemster, G.T.** (2014). Leaf development: a cellular perspective. Front Plant Sci **5,** 362.

**Kanei, M., Horiguchi, G., and Tsukaya, H.** (2012). Stable establishment of cotyledon identity during embryogenesis in Arabidopsis by ANGUSTIFOLIA3 and HANABA TARANU. Development **139,** 2436-2446.

**Karimi, M., De Meyer, B., and Hilson, P.** (2005). Modular cloning in plant cells. Trends in plant science **10,** 103-105.

**Kashi, K., Henderson, L., Bonetti, A., and Carninci, P.** (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. Biochim Biophys Acta **1859,** 3-15.

**Katsir, L., Davies, K.A., Bergmann, D.C., and Laux, T.** (2011). Peptide signaling in plant development. Curr Biol **21,** R356-364.

**Kawade, K., Horiguchi, G., and Tsukaya, H.** (2010). Non-cell-autonomously coordinated organ size regulation in leaf development. Development **137,** 4221-4227.

**Kawade, K., Horiguchi, G., Usami, T., Hirai, M.Y., and Tsukaya, H.** (2013). ANGUSTIFOLIA3 signaling coordinates proliferation between clonally distinct cells in leaves. Curr Biol **23,** 788-792.

**Kazama, T., Ichihashi, Y., Murata, S., and Tsukaya, H.** (2010). The mechanism of cell cycle arrest front progression explained by a KLUH/CYP78A5-dependent mobile growth factor in developing leaves of Arabidopsis thaliana. Plant & cell physiology **51,** 1046-1054.

**Kidd, B.N., Edgar, C.I., Kumar, K.K., Aitken, E.A., Schenk, P.M., Manners, J.M., and Kazan, K.** (2009). The Mediator Complex Subunit PFT1 Is a Key Regulator of Jasmonate-Dependent Defense in Arabidopsis. The Plant cell **21,** 2237-2252.

**Kierzkowski, D., Nakayama, N., Routier-Kierzkowska, A.L., Weber, A., Bayer, E., Schorderet, M., Reinhardt, D., Kuhlemeier, C., and Smith, R.S.** (2012). Elastic domains regulate growth and organogenesis in the plant shoot apical meristem. Science **335,** 1096-1099.

**Kim, D.H., and Sung, S.** (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. Dev Cell **40,** 302-312.

**Kim, G.T., Fujioka, S., Kozuka, T., Tax, F.E., Takatsuto, S., Yoshida, S., and Tsukaya, H.** (2005). CYP90C1 and CYP90D1 are involved in different steps in the brassinosteroid biosynthesis pathway in Arabidopsis thaliana. The Plant journal : for cell and molecular biology **41,** 710-721.

**Kim, J.H., and Kende, H.** (2004). A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in Arabidopsis. Proc Natl Acad Sci U S A **101,** 13374-13379.

**Kim, J.H., and Lee, B.H.** (2006). GROWTH-REGULATING FACTOR4 ofArabidopsis thaliana is required for development of leaves, cotyledons, and shoot apical meristem. Journal of Plant Biology **49,** 463-468.

**Kim, J.H., Choi, D., and Kende, H.** (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. The Plant journal : for cell and molecular biology **36,** 94-104.

**Kiss, T.** (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell **109,** 145-148.

**Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., and Penin, A.A.** (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. The Plant journal : for cell and molecular biology **88,** 1058-1070.

**Kojima, K., Tamura, J., Chiba, H., Fukada, K., Tsukaya, H., and Horiguchi, G.** (2017). Two Nucleolar Proteins, GDP1 and OLI2, Function As Ribosome Biogenesis Factors and Are Preferentially Involved in Promotion of Leaf Cell Proliferation without Strongly Affecting Leaf Adaxial-Abaxial Patterning in Arabidopsis thaliana. Front Plant Sci **8,** 2240.

**Koyama, T., Sato, F., and Ohme-Takagi, M.** (2017). Roles of miR319 and TCP Transcription Factors in Leaf Development. Plant physiology **175,** 874-885.

**Kozuka, T., Kong, S.G., Doi, M., Shimazaki, K., and Nagatani, A.** (2011). Tissue-autonomous promotion of palisade cell development by phototropin 2 in Arabidopsis. The Plant cell **23,** 3684-3695.

**Krchnakova, Z., Thakur, P.K., Krausova, M., Bieberstein, N., Haberman, N., Muller-McNicoll, M., and Stanek, D.** (2019). Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins. Nucleic acids research **47,** 911-928.

**Kung, J.T., Colognori, D., and Lee, J.T.** (2013). Long noncoding RNAs: past, present, and future. Genetics **193,** 651-669.

**Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A., Toyoda, T., Shinozaki, K., and Seki, M.** (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. Proc Natl Acad Sci U S A **106,** 2453-2458.

**Langlois-Meurinne, M., Gachon, C.M., and Saindrenan, P.** (2005). Pathogen-responsive expression of glycosyltransferase genes UGT73B3 and UGT73B5 is necessary for resistance to Pseudomonas syringae pv tomato in Arabidopsis. Plant physiology **139,** 1890-1901.

**Laufs, P., Peaucelle, A., Morin, H., and Traas, J.** (2004). MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. Development **131,** 4311-4322.

**Lavy, M., and Estelle, M.** (2016). Mechanisms of auxin signaling. Development **143,** 3226-3229.

**Lee, B.H., Ko, J.H., Lee, S., Lee, Y., Pak, J.H., and Kim, J.H.** (2009). The Arabidopsis GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. Plant physiology **151,** 655-668.

**Lee, B.H., Kwon, S.H., Lee, S.J., Park, S.K., Song, J.T., Lee, S., Lee, M.M., Hwang, Y.S., and Kim, J.H.** (2015). The Arabidopsis thaliana NGATHA transcription factors negatively regulate cell proliferation of lateral organs. Plant Mol Biol **89,** 529-538.

**Lee, J.T., Davidow, L.S., and Warshawsky, D.** (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nature genetics **21,** 400-404.

**Lee, Y., Choi, D., and Kende, H.** (2001). Expansins: ever-expanding numbers and functions. Curr Opin Plant Biol **4,** 527-532.

**Leivar, P., and Quail, P.H.** (2011). PIFs: pivotal components in a cellular signaling hub. Trends in plant science **16,** 19-28.

**Levine, M., and Tjian, R.** (2003). Transcription regulation and animal diversity. Nature **424,** 147-151.

**Li, L., Eichten, S.R., Shimizu, R., Petsch, K., Yeh, C.T., Wu, W., Chettoor, A.M., Givan, S.A., Cole, R.A., Fowler, J.E., Evans, M.M., Scanlon, M.J., Yu, J., Schnable, P.S., Timmermans, M.C., Springer, N.M., and Muehlbauer, G.J.** (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. Genome biology **15,** R40.

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

**Li, Q., Su, Z., Xu, X., Liu, G., Song, X., Wang, R., Sui, X., Liu, T., Chang, X., and Huang, D.** (2012). AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in cis and trans. Proc Natl Acad Sci U S A **109,** 14110-14115.

**Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N.** (2016). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev Cell **39,** 508-522.

**Li, X., Cai, W., Liu, Y., Li, H., Fu, L., Liu, Z., Xu, L., Liu, H., Xu, T., and Xiong, Y.** (2017). Differential TOR activation and cell proliferation in Arabidopsis root and shoot apexes. Proc Natl Acad Sci U S A **114,** 2765-2770.

## References

**Li, Y., Baldauf, S., Lim, E.K., and Bowles, D.J.** (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. The Journal of biological chemistry **276,** 4338-4343.

**Lim, E.K., and Bowles, D.J.** (2004). A class of plant glycosyltransferases involved in cellular homeostasis. EMBO J **23,** 2915-2922.

**Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., and Chua, N.H.** (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant cell **24,** 4333-4345.

**Liu, S.J., and Lim, D.A.** (2018). Modulating the expression of long non-coding RNAs for functional studies. EMBO Rep **19**.

**Liu, X., Li, D.Y., Zhang, D.L., Yin, D.D., Zhao, Y., Ji, C.J., Zhao, X.F., Li, X.B., He, Q., Chen, R.S., Hu, S.N., and Zhu, L.H.** (2018). A novel antisense long noncoding RNA, TWISTED LEAF, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice. New Phytol **218,** 774-788.

**Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K.** (1996). A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. Nature **379,** 66-69.

**Louro, R., Smirnova, A.S., and Verjovski-Almeida, S.** (2009). Long intronic noncoding RNA transcription: expression noise or expression choice? Genomics **93,** 291-298.

**Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P., and Lam, E.** (2013). Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. The Plant journal : for cell and molecular biology **73,** 77-90.

**Luo, J., Zhou, J.J., and Zhang, J.Z.** (2018). Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and Function. Int J Mol Sci **19**.

**Magistri, M., Faghihi, M.A., St Laurent, G., 3rd, and Wahlestedt, C.** (2012). Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. Trends in genetics : TIG **28,** 389-396.

**Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B.** (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr Biol **14,** 1035-1046.

**Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L.** (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol **13,** 95-101.

**Marchese, F.P., Raimondi, I., and Huarte, M.** (2017). The multidimensional mechanisms of long noncoding RNA function. Genome biology **18,** 206.

**Marowa, P., Ding, A., and Kong, Y.** (2016). Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep **35,** 949-965.

**Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., and Dean, C.** (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. Molecular cell **54,** 156-165.

**Mathur, S., Vyas, S., Kapoor, S., and Tyagi, A.K.** (2011). The Mediator Complex in Plants: Structure, Phylogeny, and Expression Profiling of Representative Genes in a Dicot (Arabidopsis) and a Monocot (Rice) during Reproduction and Abiotic Stress. Plant physiology **157,** 1609-1627.

**Matsui, A., Iida, K., Tanaka, M., Yamaguchi, K., Mizuhashi, K., Kim, J.M., Takahashi, S., Kobayashi, N., Shigenobu, S., Shinozaki, K., and Seki, M.** (2017). Novel Stress-Inducible Antisense RNAs of Protein-Coding Loci Are Synthesized by RNA-Dependent RNA Polymerase. Plant physiology **175,** 457-472.

**Mattick, J.S.** (2005). The functional genomics of noncoding RNA. Science **309,** 1527-1528.

**Mattick, J.S.** (2009). The genetic signatures of noncoding RNAs. PLoS genetics **5,** e1000459.

**McQueen-Mason, S., Durachko, D.M., and Cosgrove, D.J.** (1992). Two endogenous proteins that induce cell wall extension in plants. The Plant cell **4,** 1425-1433.

**McSteen, P., and Zhao, Y.** (2008). Plant hormones and signaling: common themes and new developments. Dev Cell **14,** 467-473.

**Meijering, E., Jacob, M., Sarria, J.C., Steiner, P., Hirling, H., and Unser, M.** (2004). Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry. Part A : the journal of the International Society for Analytical Cytology **58,** 167-176.

**Milani, P., Gholamirad, M., Traas, J., Arnéodo, A., Boudaoud, A., Argoul, F., and Hamant, O.** (2011). In   vivo analysis of local wall stiffness at the shoot apical meristem in Arabidopsis using atomic force microscopy. The Plant Journal **67,** 1116-1123.

**Mizukami, Y., and Fischer, R.L.** (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proc Natl Acad Sci U S A **97,** 942-947.

**Moreau, M., Azzopardi, M., Clement, G., Dobrenel, T., Marchive, C., Renne, C., Martin-Magniette, M.L., Taconnat, L., Renou, J.P., Robaglia, C., and Meyer, C.** (2012). Mutations in the Arabidopsis homolog of LST8/GbetaL, a partner of the target of Rapamycin kinase, impair plant growth, flowering, and metabolic adaptation to long days. The Plant cell **24,** 463-481.

**Moriya, H.** (2015). Quantitative nature of overexpression experiments. Mol Biol Cell **26,** 3932-3939.

**Morris, K.V., and Mattick, J.S.** (2014). The rise of regulatory RNA. Nature reviews. Genetics **15,** 423-437.

**Nakata, M., and Okada, K.** (2013). The Leaf Adaxial-Abaxial Boundary and Lamina Growth. Plants (Basel) **2,** 174-202.

**Narita, N.N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J., and Tsukaya, H.** (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. The Plant journal : for cell and molecular biology **38,** 699-713.

**Nath, U., Crawford, B.C., Carpenter, R., and Coen, E.** (2003). Genetic control of surface curvature. Science **299,** 1404-1407.

**Nelissen, H., Rymen, B., Coppens, F., Dhondt, S., Fiorani, F., and Beemster, G.T.** (2013). Kinematic analysis of cell division in leaves of mono- and dicotyledonous species: a basis for understanding growth and developing refined molecular sampling strategies. Methods in molecular biology **959,** 247-264.

**Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., Cannon, S.C., Houser, S.R., Bassel-Duby, R., and Olson, E.N.** (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science (New York, N.Y.) **351,** 271-275.

**Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,**

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y., Consortium, F., I, R.G.E.R.G.P., and Team, I.I. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature **420,** 563-573.

Omidbakhshfard, M.A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-Regulating Factors (GRFs): A Small Transcription Factor Family with Important Functions in Plant Biology. Molecular plant **8,** 998-1010.

Osmani, S.A., Bak, S., Imberty, A., Olsen, C.E., and Moller, B.L. (2008). Catalytic key amino acids and UDP-sugar donor specificity of a plant glucuronosyltransferase, UGT94B1: molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant physiology **148,** 1295-1308.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nature reviews. Genetics **12,** 87-98.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature **425,** 257-263.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Frontiers in genetics **6,** 2.

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

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O., Hughes, J.R., Hardison, R.C., Blobel, G.A., and Weiss, M.J. (2016). Unlinking an lncRNA from Its Associated cis Element. Molecular cell **62,** 104-110.

Paralkar, V.R., Mishra, T., Luan, J., Yao, Y., Kossenkov, A.V., Anderson, S.M., Dunagin, M., Pimkin, M., Gore, M., Sun, D., Konuthula, N., Raj, A., An, X., Mohandas, N., Bodine, D.M., Hardison, R.C., and Weiss, M.J. (2014). Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. Blood **123,** 1927-1937.

Pelaz, S., Tapia-Lopez, R., Alvarez-Buylla, E.R., and Yanofsky, M.F. (2001). Conversion of leaves into petals in Arabidopsis. Curr Biol **11,** 182-184.

Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. Nature reviews. Genetics **14,** 880-893.

Perrot-Rechenmann, C. (2010). Cellular responses to auxin: division versus expansion. Cold Spring Harbor perspectives in biology **2,** a001446.

Planas-Riverola, A., Gupta, A., Betegon-Putze, I., Bosch, N., Ibanes, M., and Cano-Delgado, A.I. (2019). Brassinosteroid signaling in plant development and adaptation to stress. Development **146**.

Poethig, R.S., and Sussex, I.M. (1985). The cellular parameters of leaf development in tobacco: a clonal analysis. Planta **165,** 170-184.

Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., Schuhmacher, R., Krska, R., Kuchler, K., Glossl, J., Luschnig, C., and Adam, G. (2003). Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from Arabidopsis thaliana. The Journal of biological chemistry **278,** 47905-47914.

**Poppenberger, B., Fujioka, S., Soeno, K., George, G.L., Vaistij, F.E., Hiranuma, S., Seto, H., Takatsuto, S., Adam, G., Yoshida, S., and Bowles, D.** (2005). The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids. Proc Natl Acad Sci U S A **102,** 15253-15258.

**Prelich, G.** (2012). Gene overexpression: uses, mechanisms, and interpretation. Genetics **190,** 841-854.

**Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., Cao, X., Jing, X., Wang, X., Siddiqui, J., Wei, J.T., Robinson, D., Iyer, H.K., Palanisamy, N., Maher, C.A., and Chinnaiyan, A.M.** (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nature biotechnology **29,** 742-749.

**Quek, X.C., Thomson, D.W., Maag, J.L., Bartonicek, N., Signal, B., Clark, M.B., Gloss, B.S., and Dinger, M.E.** (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic acids research **43,** D168-173.

**Reams, A.B., and Roth, J.R.** (2015). Mechanisms of gene duplication and amplification. Cold Spring Harbor perspectives in biology **7,** a016592.

**Reinhardt, D., Frenz, M., Mandel, T., and Kuhlemeier, C.** (2005). Microsurgical and laser ablation analysis of leaf positioning and dorsoventral patterning in tomato. Development **132,** 15-26.

**Ren, T., Weraduwage, S.M., and Sharkey, T.D.** (2019). Prospects for enhancing leaf photosynthetic capacity by manipulating mesophyll cell morphology. J Exp Bot **70,** 1153-1165.

**Reyes, A., and Huber, W.** (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic acids research **46,** 582-592.

**Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., and Brandle, J.** (2005). Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana. The Plant journal : for cell and molecular biology **41,** 56-67.

**Rodriguez, R.E., Mecchia, M.A., Debernardi, J.M., Schommer, C., Weigel, D., and Palatnik, J.F.** (2010). Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. Development **137,** 103-112.

**Rosa, S., Duncan, S., and Dean, C.** (2016). Mutually exclusive sense-antisense transcription at FLC facilitates environmentally induced gene repression. Nature communications **7**.

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

**Saini, K., Markakis, M.N., Zdanio, M., Balcerowicz, D.M., Beeckman, T., De Veylder, L., Prinsen, E., Beemster, G.T.S., and Vissenberg, K.** (2017). Alteration in Auxin Homeostasis and Signaling by Overexpression Of PINOID Kinase Causes Leaf Growth Defects in Arabidopsis thaliana. Front Plant Sci **8,** 1009.

**Santos-Pereira, J.M., and Aguilera, A.** (2015). R loops: new modulators of genome dynamics and function. Nature reviews. Genetics **16,** 583-597.

**Sapala, A., Runions, A., and Smith, R.S.** (2019). Mechanics, geometry and genetics of epidermal cell shape regulation: different pieces of the same puzzle. Curr Opin Plant Biol **47,** 1-8.

**Schadt, E.E., Turner, S., and Kasarskis, A.** (2010). A window into third-generation sequencing. Human molecular genetics **19,** R227-240.

**Schneider, C.A., Rasband, W.S., and Eliceiri, K.W.** (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods **9,** 671-675.

**Schommer, C., Debernardi, J.M., Bresso, E.G., Rodriguez, R.E., and Palatnik, J.F.** (2014). Repression of cell proliferation by miR319-regulated TCP4. Molecular plant **7,** 1533-1544.

**Schweiger, W., Boddu, J., Shin, S., Poppenberger, B., Berthiller, F., Lemmens, M., Muehlbauer, G.J., and Adam, G.** (2010). Validation of a candidate deoxynivalenol-inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. Mol Plant Microbe Interact **23,** 977-986.

**Seo, J.S., Diloknawarit, P., Park, B.S., and Chua, N.H.** (2019). ELF18-INDUCED LONG NONCODING RNA 1 evicts fibrillarin from mediator subunit to enhance PATHOGENESIS-RELATED GENE 1 (PR1) expression. New Phytol **221,** 2067-2079.

**Seo, J.S., Sun, H.X., Park, B.S., Huang, C.H., Yeh, S.D., Jung, C., and Chua, N.H.** (2017). ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in Arabidopsis. The Plant cell **29,** 1024-1038.

**Severing, E., Faino, L., Jamge, S., Busscher, M., Kuijer-Zhang, Y., Bellinazzo, F., Busscher-Lange, J., Fernandez, V., Angenent, G.C., Immink, R.G.H., and Pajoro, A.** (2018). Arabidopsis thaliana ambient temperature responsive lncRNAs. BMC plant biology **18,** 145.

**Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L.** (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nature methods **12,** 664-670.

**Shi, J., Dong, J., Xue, J., Wang, H., Yang, Z., Jiao, Y., Xu, L., and Huang, H.** (2017). Model for the role of auxin polar transport in patterning of the leaf adaxial-abaxial axis. The Plant journal : for cell and molecular biology **92,** 469-480.

**Signor, S.A., and Nuzhdin, S.V.** (2018). The Evolution of Gene Expression in cis and trans. Trends in genetics : TIG **34,** 532-544.

**Smith, L.G., and Hake, S.** (1992). The Initiation and Determination of Leaves. The Plant cell **4,** 1017-1027.

**SNOW, M., and SNOW, R.** (1959). THE DORSIVENTRALITY OF LEAF PRIMORDIA. New Phytol **58,** 188-207.

**Steeves, T.A., and Sussex, I.M.** (1989). Patterns in Plant Development. (Cambridge: Cambridge University Press).

**Su, W.-Y., Li, J.-T., Cui, Y., Hong, J., Du, W., Wang, Y.-C., Lin, Y.-W., Xiong, H., Wang, J.-L., Kong, X., Gao, Q.-Y., Wei, L.-P., and Fang, J.-Y.** (2012). Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. Cell Research **22,** 1374-1389.

**Sun, Q., Hao, Q., and Prasanth, K.V.** (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. Trends in genetics : TIG **34,** 142-157.

**Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C.** (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science **340,** 619-621.

**Sussex, I.M.** (1951). Experiments on the Cause of Dorsiventrality in Leaves. Nature **167,** 651-652.

**Suzuki, M., Murayama, E., Inoue, H., Ozaki, N., Tohse, H., Kogure, T., and Nagasawa, H.** (2004). Characterization of Prismalin-14, a novel matrix protein from the prismatic layer of the Japanese pearl oyster (Pinctada fucata). The Biochemical journal **382,** 205-213.

**Swiezewski, S., Liu, F., Magusin, A., and Dean, C.** (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature **462,** 799-802.

**Taylor, C.B.** (1997). Promoter Fusion Analysis: An Insufficient Measure of Gene Expression. The Plant cell **9,** 273-275.

**Terryn, N., and Rouze, P.** (2000). The sense of naturally transcribed antisense RNAs in plants. Trends in plant science **5,** 394-396.

**Tomizawa, J., Itoh, T., Selzer, G., and Som, T.** (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. Proc Natl Acad Sci U S A **78,** 1421-1425.

**Travella, S., and Keller, B.** (2009). Down-regulation of gene expression by RNA-induced gene silencing. Methods in molecular biology **478,** 185-199.

**Triller, S.** (2019). Identification of UGT73C6 and lncNATs-UGT73C6 transcription initiation and termination sites and analysis of lncNATs role in leaf growth. In Department of Molecular Signal Processing,  Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany (Halle (Saale), Germany: Martin-Luther-Universität Halle-Wittenberg).

**Tsuge, T., Tsukaya, H., and Uchimiya, H.** (1996). Two independent and polarized processes of cell elongation regulate leaf blade expansion in Arabidopsis thaliana (L.) Heynh. Development **122,** 1589-1600.

**Tsukaya, H.** (2002a). Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. Int Rev Cytol **217,** 1-39.

**Tsukaya, H.** (2002b). Leaf development. Arabidopsis Book **1,** e0072.

**Tsukaya, H.** (2003). Organ shape and size: a lesson from studies of leaf morphogenesis. Curr Opin Plant Biol **6,** 57-62.

**Tsukaya, H.** (2013a). RTFL/DVL peptide family regulation of shoot development. Pegtide Signals in Plants.

**Tsukaya, H.** (2013b). Leaf development. Arabidopsis Book **11,** e0163.

**Tsutsui, H., and Higashiyama, T.** (2017). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in Arabidopsis thaliana. Plant & cell physiology **58,** 46-56.

**Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W.S., Karpilow, J., and Khvorova, A.** (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. RNA **11,** 674-682.

**Vert, G., Nemhauser, J.L., Geldner, N., Hong, F., and Chory, J.** (2005). Molecular mechanisms of steroid hormone signaling in plants. Annu Rev Cell Dev Biol **21,** 177-201.

**Vogt, T., and Jones, P.** (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends in plant science **5,** 380-386.

**Vroemen, C.W., Mordhorst, A.P., Albrecht, C., Kwaaitaal, M.A., and de Vries, S.C.** (2003). The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis. The Plant cell **15,** 1563-1577.

**Wang, H.-F., Feng, L., and Niu, D.-K.** (2007). Relationship between mRNA stability and intron presence. Biochemical and Biophysical Research Communications **354,** 203-208.

**Wang, H., Chua, N.H., and Wang, X.J.** (2006). Prediction of trans-antisense transcripts in Arabidopsis thaliana. Genome biology **7,** R92.

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

**Wang, H.L.V., and Chekanova, J.A.** (2017a). Long Noncoding RNAs in Plants. Long Non Coding Rna Biology **1008,** 133-154.

**Wang, H.V., and Chekanova, J.A.** (2017b). Long Noncoding RNAs in Plants. Advances in experimental medicine and biology **1008,** 133-154.

**Wang, J., Tao, F., Marowsky, N.C., and Fan, C.** (2016). Evolutionary Fates and Dynamic Functionalization of Young Duplicate Genes in Arabidopsis Genomes. Plant physiology **172,** 427-440.

**Wang, J.J., and Guo, H.S.** (2015). Cleavage of INDOLE-3-ACETIC ACID INDUCIBLE28 mRNA by microRNA847 upregulates auxin signaling to modulate cell proliferation and lateral organ growth in Arabidopsis. The Plant cell **27,** 574-590.

**Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z., and Wu, W.H.** (2008). Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. Plant physiology **148,** 1201-1211.

**Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.S., Gautheret, D., and Morillon, A.** (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Molecular cell **61,** 379-392.

**Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S.** (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell **135,** 635-648.

**Williams, T., and Fried, M.** (1986). A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature **322,** 275-279.

**Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G.** (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science **309,** 1570-1573.

**Wunderlich, M., Gross-Hardt, R., and 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,** 541-550.

**Xiong, Y., and Sheen, J.** (2012). Rapamycin and glucose-target of rapamycin (TOR) protein signaling in plants. The Journal of biological chemistry **287,** 2836-2842.

**Xu, C., Park, J.-K., and Zhang, J.** (2019). Evidence that alternative transcriptional initiation is largely nonadaptive. PLOS Biology **17,** e3000197.

**Xu, L., Xu, Y., Dong, A., Sun, Y., Pi, L., Xu, Y., and Huang, H.** (2003). Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity. Development **130,** 4097-4107.

**Xue, Z., Hennelly, S., Doyle, B., Gulati, A.A., Novikova, I.V., Sanbonmatsu, K.Y., and Boyer, L.A.** (2016). A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. Molecular cell **64,** 37-50.

**Yatusevich, R., Fedak, H., Ciesielski, A., Krzyczmonik, K., Kulik, A., Dobrowolska, G., and Swiezewski, S.** (2017). Antisense transcription represses Arabidopsis seed dormancy QTL DOG1 to regulate drought tolerance. EMBO Rep **18,** 2186-2196.

**Yuan, J., Zhang, Y., Dong, J., Sun, Y., Lim, B.L., Liu, D., and Lu, Z.J.** (2016). Systematic characterization of novel lncRNAs responding to phosphate starvation in Arabidopsis thaliana. BMC genomics **17,** 655.

**Zhan, S., and Lukens, L.** (2013). Protein-Coding cis-Natural Antisense Transcripts Have High and Broad Expression in Arabidopsis. Plant physiology **161,** 2171.

**Zhang, Y., Persson, S., and Giavalisco, P.** (2013). Differential regulation of carbon partitioning by the central growth regulator target of rapamycin (TOR). Molecular plant **6,** 1731-1733.

**Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S., and Chen, Y.Q.** (2014). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome biology **15,** 512.

**Zhang, Z., Zheng, Y., Ham, B.-K., Zhang, S., Fei, Z., and Lucas, W.J.** (2019). Plant lncRNAs are enriched in and move systemically through the phloem in response to phosphate deficiency. Journal of Integrative Plant Biology **61,** 492-508.

**Zhao, D., Yang, M., Solava, J., and Ma, H.** (1999). The ASK1 gene regulates development and interacts with the UFO gene to control floral organ identity in Arabidopsis. Dev Genet **25,** 209-223.

**Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y.** (2018). Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nature communications **9,** 5056.

**Zhiponova, M.K., Vanhoutte, I., Boudolf, V., Betti, C., Dhondt, S., Coppens, F., Mylle, E., Maes, S., Gonzalez-Garcia, M.P., Cano-Delgado, A.I., Inze, D., Beemster, G.T., De Veylder, L., and Russinova, E.** (2013). Brassinosteroid production and signaling differentially control cell division and expansion in the leaf. New Phytol **197,** 490-502.

**Zhu, J.Y., Sae-Seaw, J., and Wang, Z.Y.** (2013). Brassinosteroid signalling. Development **140,** 1615-1620.

**Zhu, Q.H., Stephen, S., Taylor, J., Helliwell, C.A., and Wang, M.B.** (2014). Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. New Phytol **201,** 574-584.

## Appendix

### Supplementary Figures

| | 5 DAS | 14 DAS | 6th Leaf 25 DAS |
|---|---|---|---|
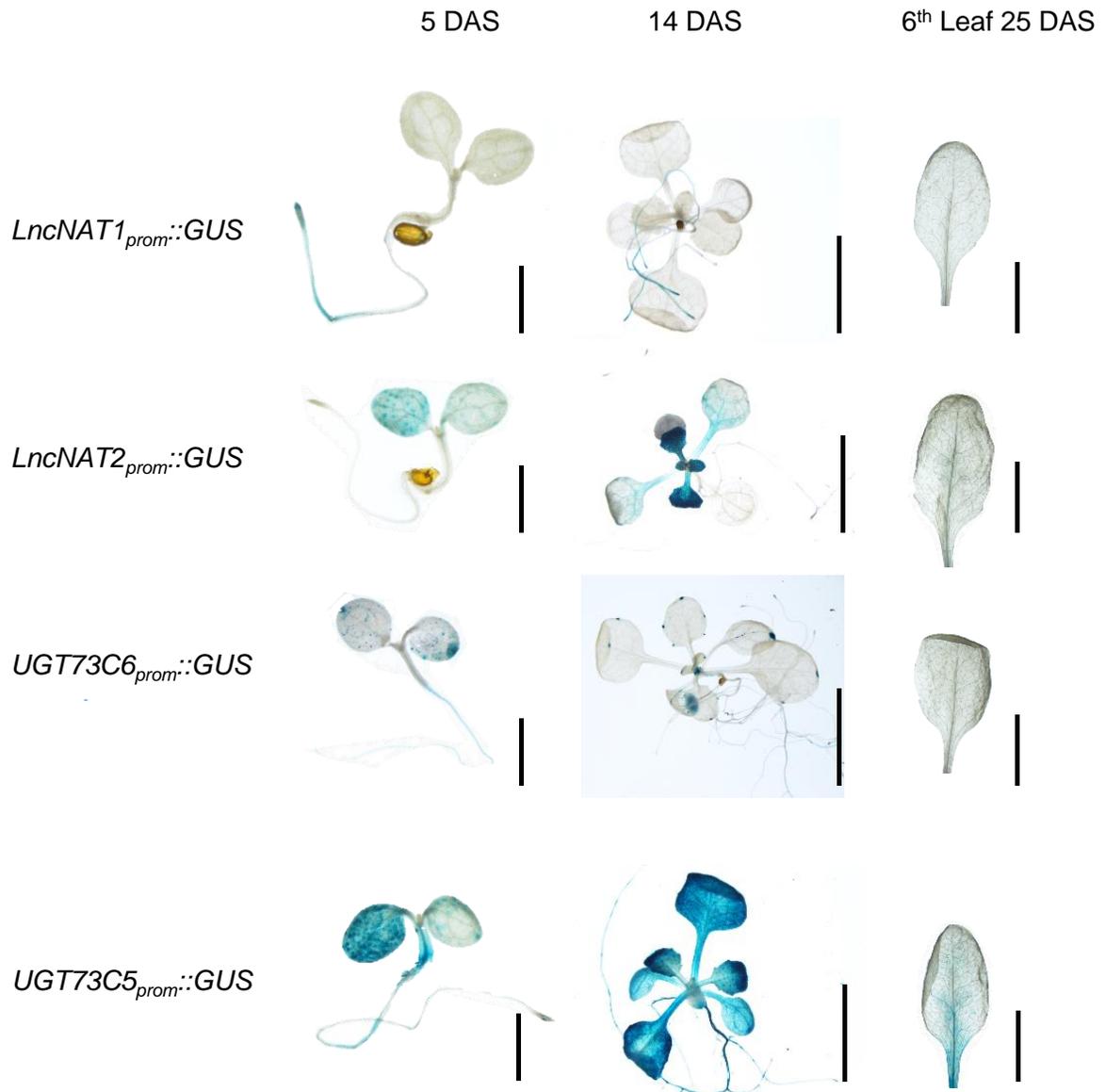


Supplementary Figure 1: **LncNATs-UGT73C6, UGT73C5 and UGT73C6 are expressed in developmentally controlled manner in _A. thaliana_**. A) Time course _prom::GUS_ expression analysis for _lncNAT1, lncNAT2, UGT73C6_ and _UGT73C5_ at 5, 14 and 25 DAS. Plant material was collected at fixed time points at all indicated developmental stages and incubated in GUS solution over a period of 12 hours. Shown are representative seedlings and leaves from one of 3 independent transgenic lines, that were generated in _Col-0_ background. Data are representative of one of at least 3 independent experiments with similar results. Scale bar 5 mm.

A)



B)
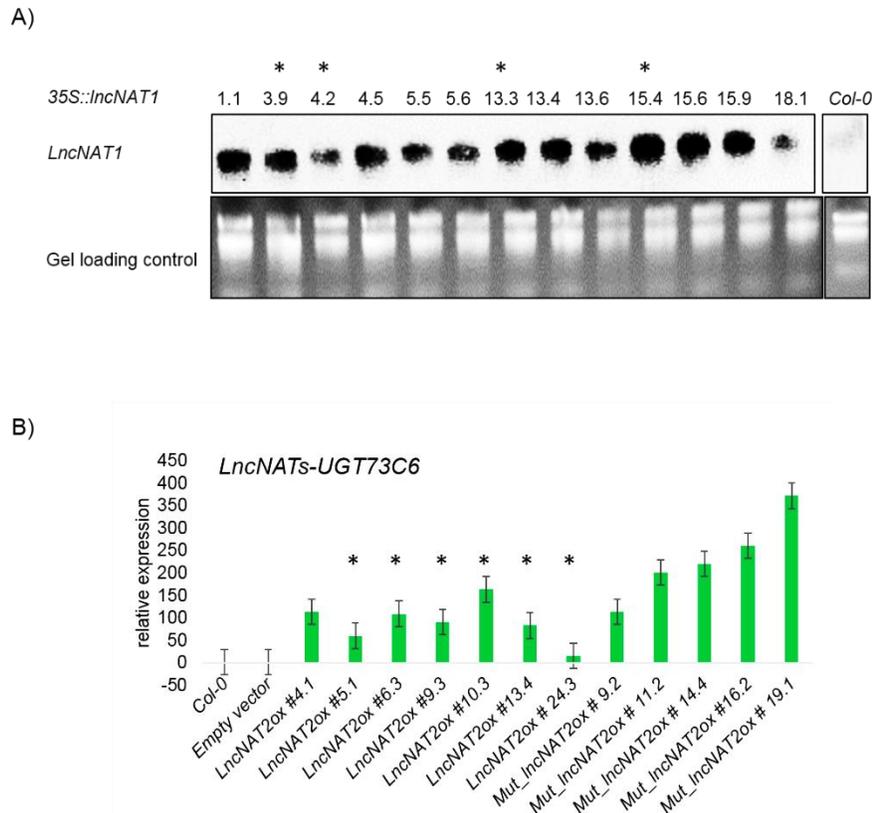


Supplementary Figure 2: **Levels of overexpression in *35S::lncNAT1* and *35S::lncNAT2* lines**. A) Levels of *lncNAT1* detected by northern blot using a strand specific radioactively labelled riboprobe (upper panel). Numbers represent independent T$_3$ homozygous lines. Asterisks indicate lines that were chosen for studying the effects of *lncNAT1* overexpression. RNA denaturing gel picture shows loading control (Bottom panel). 5µg of total RNA was loaded onto the gel and blots were visualized using Typhoon FLA 7000 phosphoimager. B) qPCR bar graphs showing relative transcripts levels of *lncNATs-UGT73C6* in *lncNAT2* and *mut_lncNAT2* overexpression lines. Error bars represent standard error.
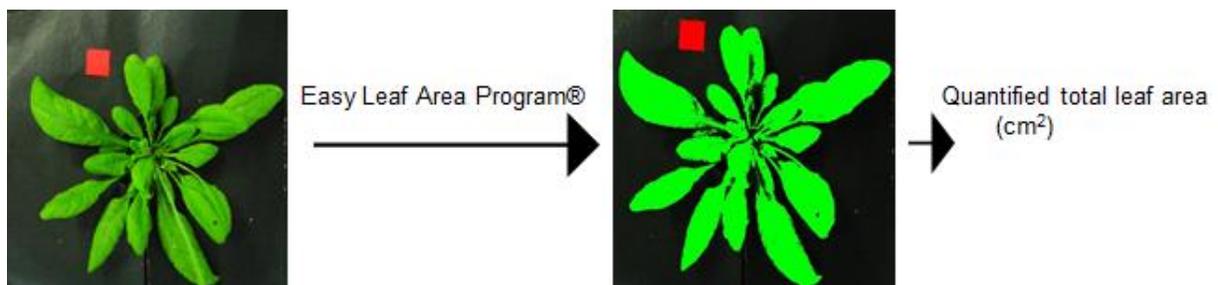


Supplementary Figure 3: **Representative scheme showing procedural steps in quantification of complete rosette images by Easy leaf area®**. Original (left) and processed image (right). Rosettes were photographed to obtain images in .jpg (joint photographic group) format using ordinary single-lens reflex (SLR) camera at manually fixed pixel settings. Digital rosette images were loaded onto the program to acquire quantified value for green pixels. 1 cm$^2$ red squire object is used as scale according to the program manual (Easlon *et. al*, 2014). Shown are picture and processed rosette image from a 5 week old plant.
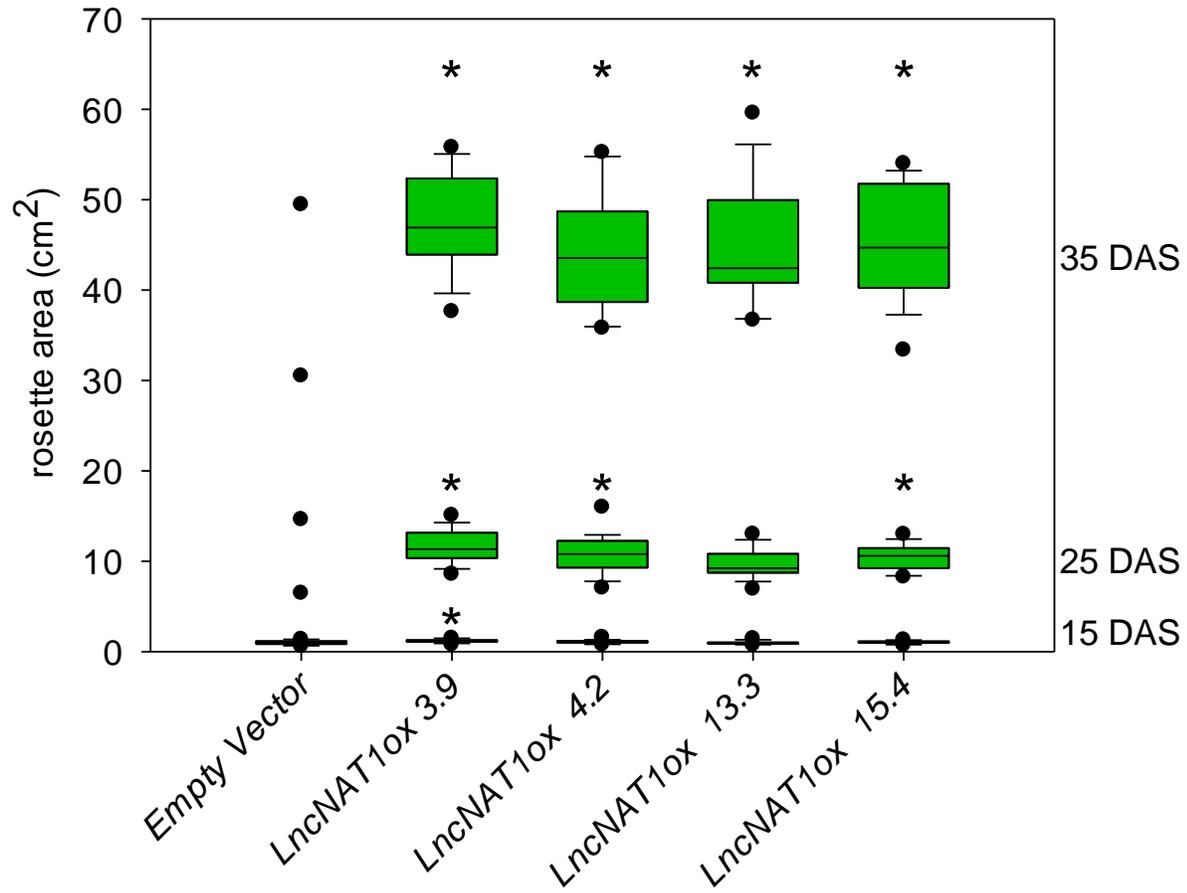
Supplementary Figure 4: **Effects due to *lncNAT1* overexpression are stronger after 25 DAS**. *LncNAT1* overexpression effects over leaf area at 15, 25 and 35 DAS. Plants were grown in controlled conditions in growth cabinet as mentioned in text. Individual *lncNAT1* overexpression lines are compared with plants transformed with the empty vector. Each group contains 20 individual plants. Asterisks indicate statistical significance compared to control plants for respective time point. * <0.05, (one-way ANOVA).
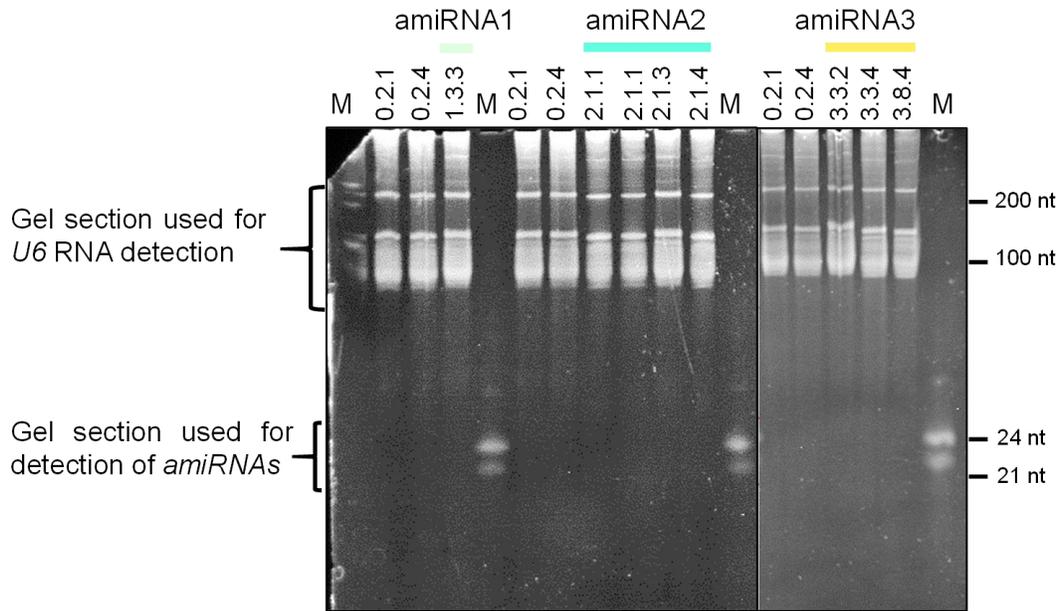
Supplementary Figure 5: **17% denaturing polyacrylamide gel electrophoresis (PAGE) for low molecular weight RNAs separation.** Total RNA from pool of homozygous $T_3$ plants was extracted by phenolic extraction and 30 μg of RNA was loaded per lane. Gel portions that were used for *U6* (blue dotted lines) and mature *amiRNAs* detection (red dotted lines) in northern blot are indicated. Empty vector and *amiRNA* lines (colour code) are marked at the top of each lane. M: Marker.
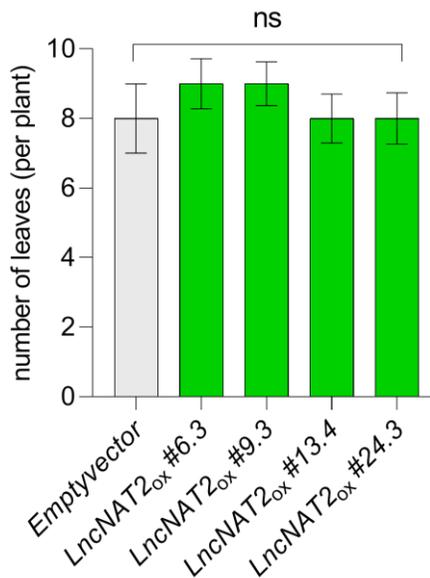


Supplementary Figure 6: ***LncNAT2* overexpression does not affect leaf number in *A. thaliana.*** Histogram showing median number of rosette leaves per plant at 25 DAS for *lncNAT2* overexpression lines and plants transformed with the empty vector. Error bars represent standard deviation. Each group contains 20 individual plants. There are no statistically significant differences among *lncNAT2* overexpression lines and control plants transformed with the empty vector. $p > 0.05$, one-way ANOVA.
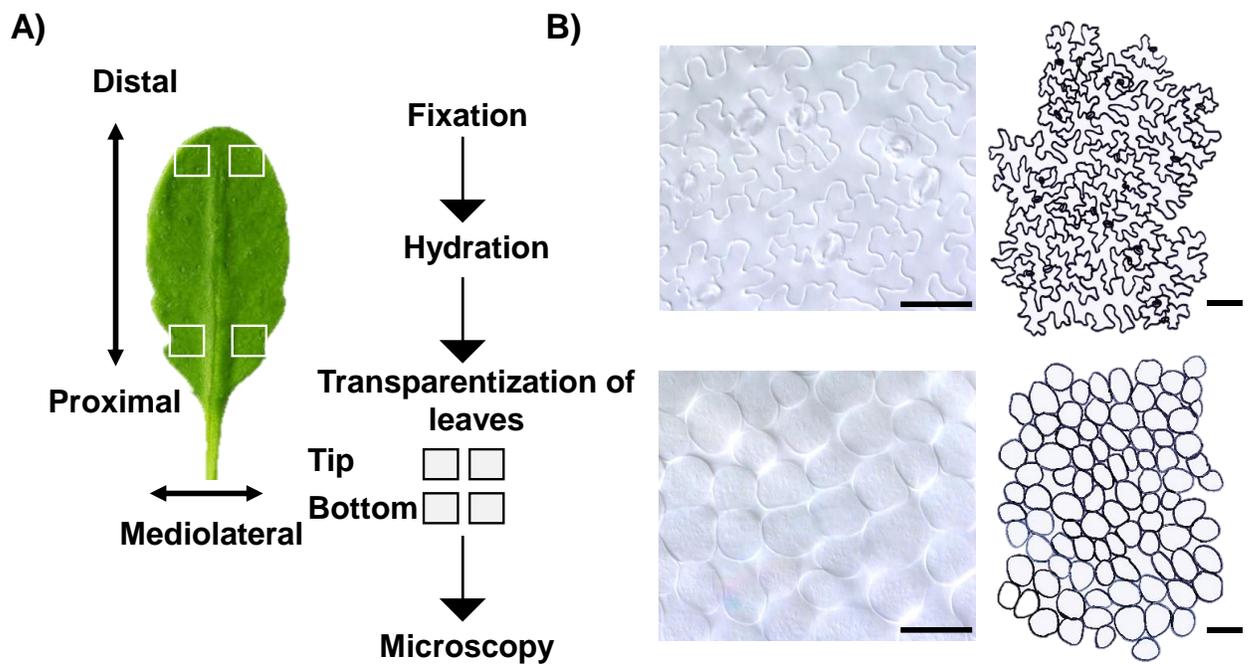
Supplementary Figure 7: **Procedural steps for DIC microscopy from leaves**. (A) Intact leaf is made transparent using Hoyer's medium and/or D-lactic acid followed by sectioning and imaging in highlighted region both at proximal and distal leaf zones. (B) Adaxial epidermis (top left) and palisade mesophyll cells (bottom left) as seen under microscope were hand drawn over transparent sheets and further quantified for cell size using image J as reported by Nelissen et al., 2013. Scale bar 100 um.

A)



B)



C)



D)
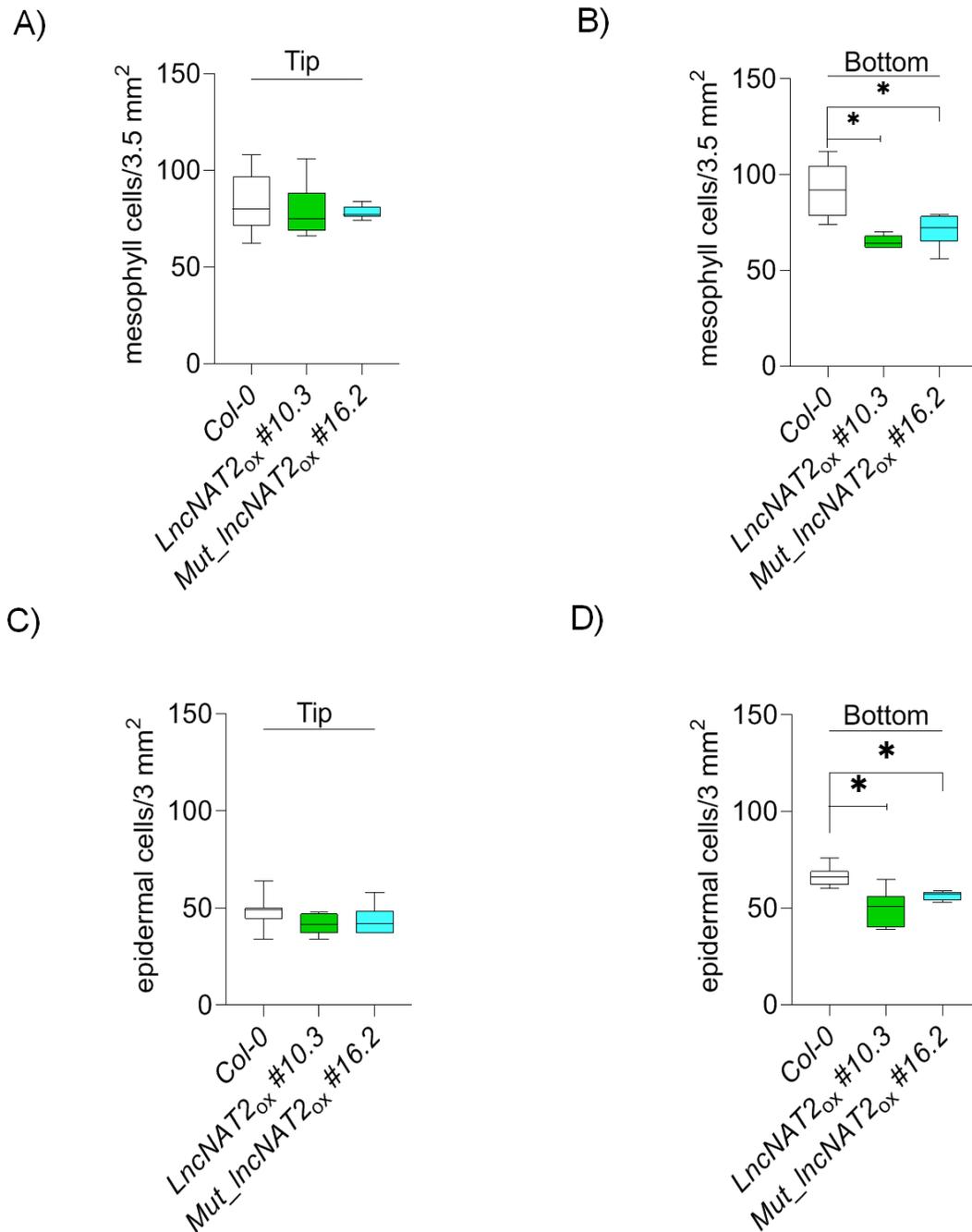


Supplementary Figure 8: *LncNAT2* **overexpression effects over the number of mesophyll and epidermal cells**. A to D) Box plots showing number of mesophyll (upper panel) and epidermal cells (bottom panel) in tip (left) and bottom (right) regions of 6th leaves for *Col-0*, *lncNAT2ox #10.3* and *Mut_lncNAT2ox #16.6*. Guard cells were excluded in the calculation of numbers of adaxial epidermal cells. 7 to 12 images from 2 leaves were included in analysis. 6th leaves were collected from 35 DAS plants. * $p < 0.05$ (one-way ANOVA).
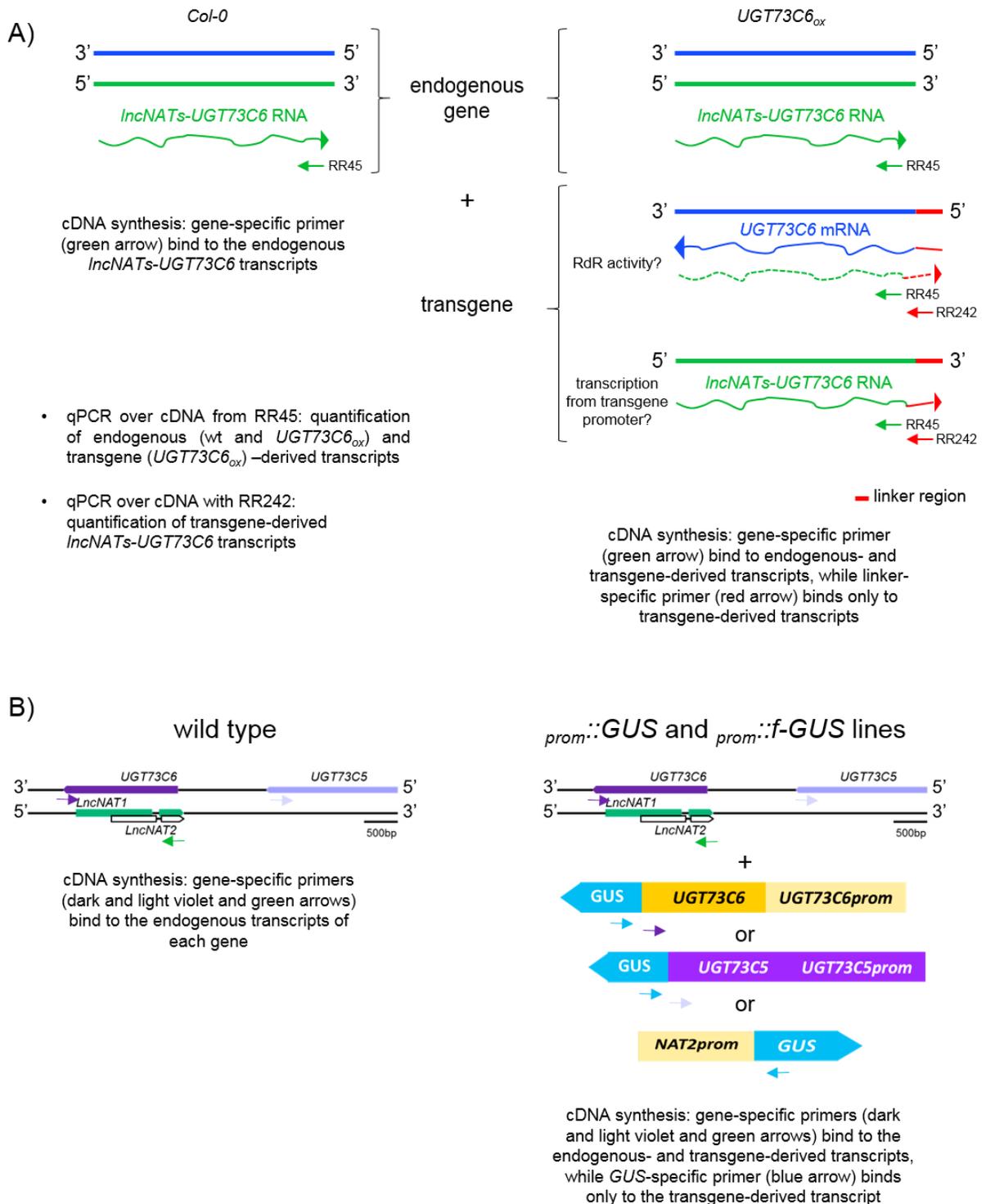
Supplementary Figure 9: Experimental scheme for synthesis of cDNA and qPCR analysis to differentiate the amplification of *lncNATs-UGT73C6, UGT73C6* and *UGT73C5* from endogenous and transgene copies of the gene in wild type and transgenic *UGT73C6ox* (A) and reporter (B) lines, respectively. Schematic shown for genomic locus for *GUS* lines is not in scale.

**A)**



5 DAS seedlings
grown over ½ MS
agar plates

48 hours adaptation in liquid
media, continuous light

Brassinolide
(1 µM 24-Epi BL)

RNA extraction, cDNA
preparation, qPCR analysis

Image Credit: Yvon Jaillais

**B)**
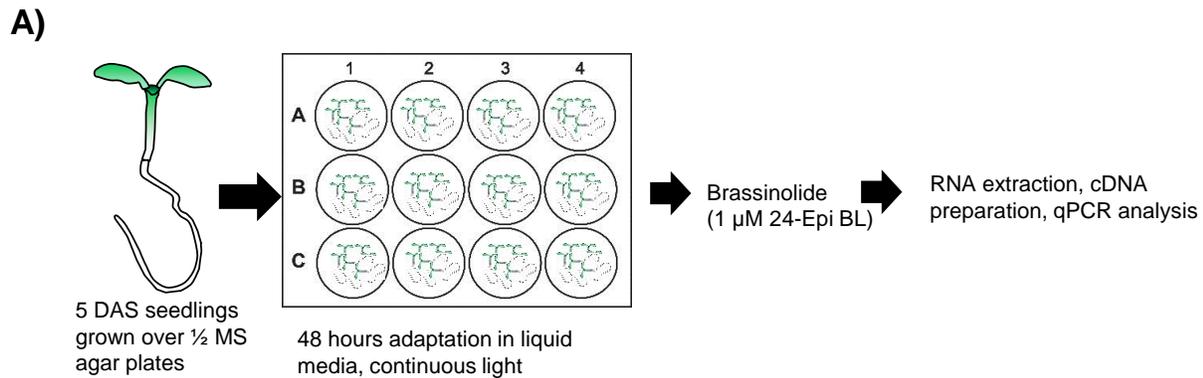


Supplementary Figure 10: **24-Epi-BL treatment results in downregulation of BR biosynthetic genes.**
A) Scheme of BR treatment assay. B) qPCR quantification of *BRASSINOSTEROID-6-OXIDASE 2 BR6ox2*
(left) and *CYTOCHROME P450 90B1* (*DWF4*) (right) in 24-Epi-BL and control (DMSO) treated *Col-0*
seedlings. Shown are combined results from 2 (10, 30 and 60 min for *BR6ox2*) or 3 independent
experiments with 6 or 9 biological replicates, respectively. ±SD, * p<0.05, Student's t-test.

A)



5 DAS seedlings grown over ½ MS agar plates

Image credit: Yvon Jaillais

48 hours adaptation in liquid media, 16/8 hour light dark cycle

DON treatment (16.9 µM DON in DMSO)

RNA extraction, cDNA preparation & qPCR

GUS staining

B)

$LncNAT1_{prom}::GUS$

Mock ($C_2H_5OH$) | +DON (deoxynivalenol)



GUS

DON(-)
DON(+)

relative expression ($\log_{10}$)

*
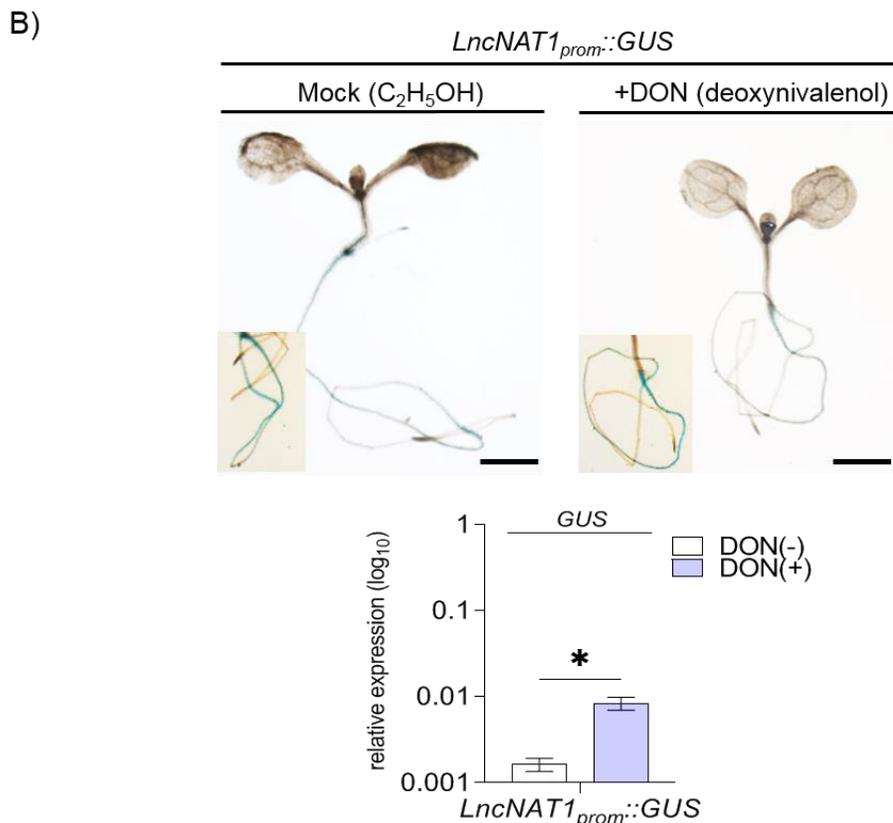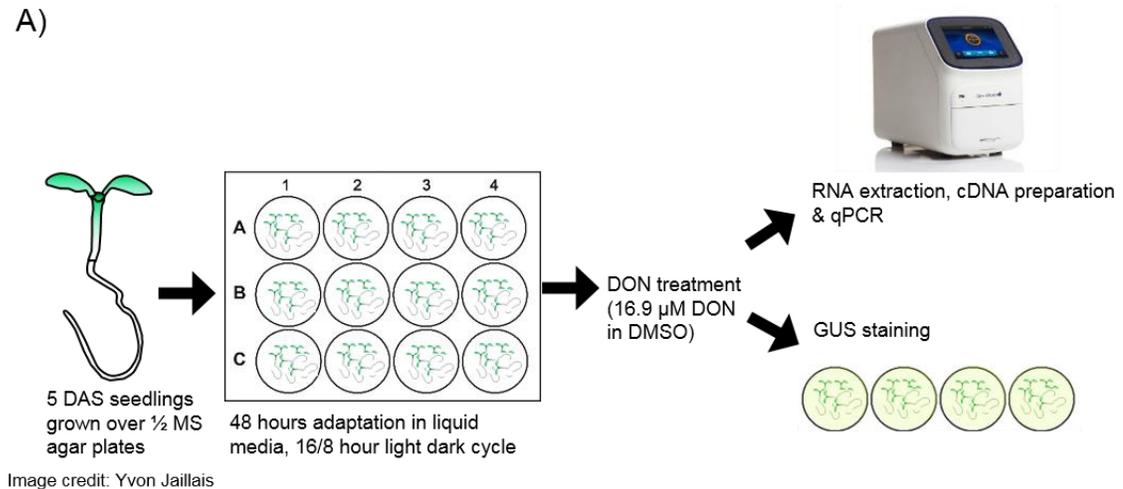
$LncNAT1_{prom}::GUS$
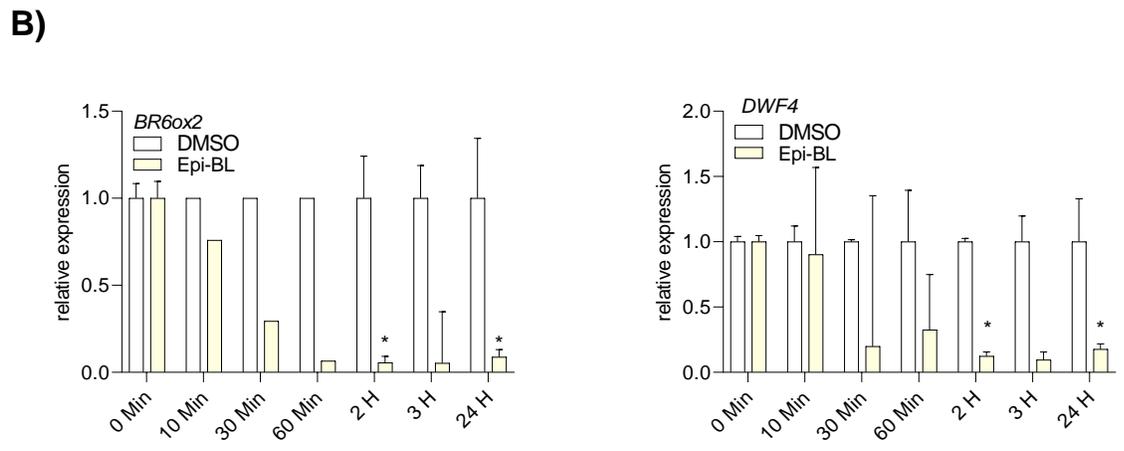
Supplementary Figure 11: **DON treatment.** A) Experimental strategy for the DON response analysis. 7 DAS old seedlings of *Col-0* (WT) and *prom::GUS* and *prom::f-GUS* lines of *lncNATs-UGT73C6*, *UGT73C6* and *UGT73C5* were treated with DON or ethanol (control) as shown in the scheme. Apart from qPCR analysis, a portion of seedlings from the assay were separately subjected to GUS staining and visualization. B) Representative pictures for mock and DON treatment are shown for *lncNAT1_{prom}::GUS*. Brightened image in inset shows contrasted roots (upper panel). qPCR measurements of *GUS* transcripts in *lncNAT1_{prom}::GUS* line after DON treatment (bottom panel). Error bars are ±SD, * p<0.05. Results from one experiment with 3 biological replicates, Student's t-test. Scale bar 1 mm.
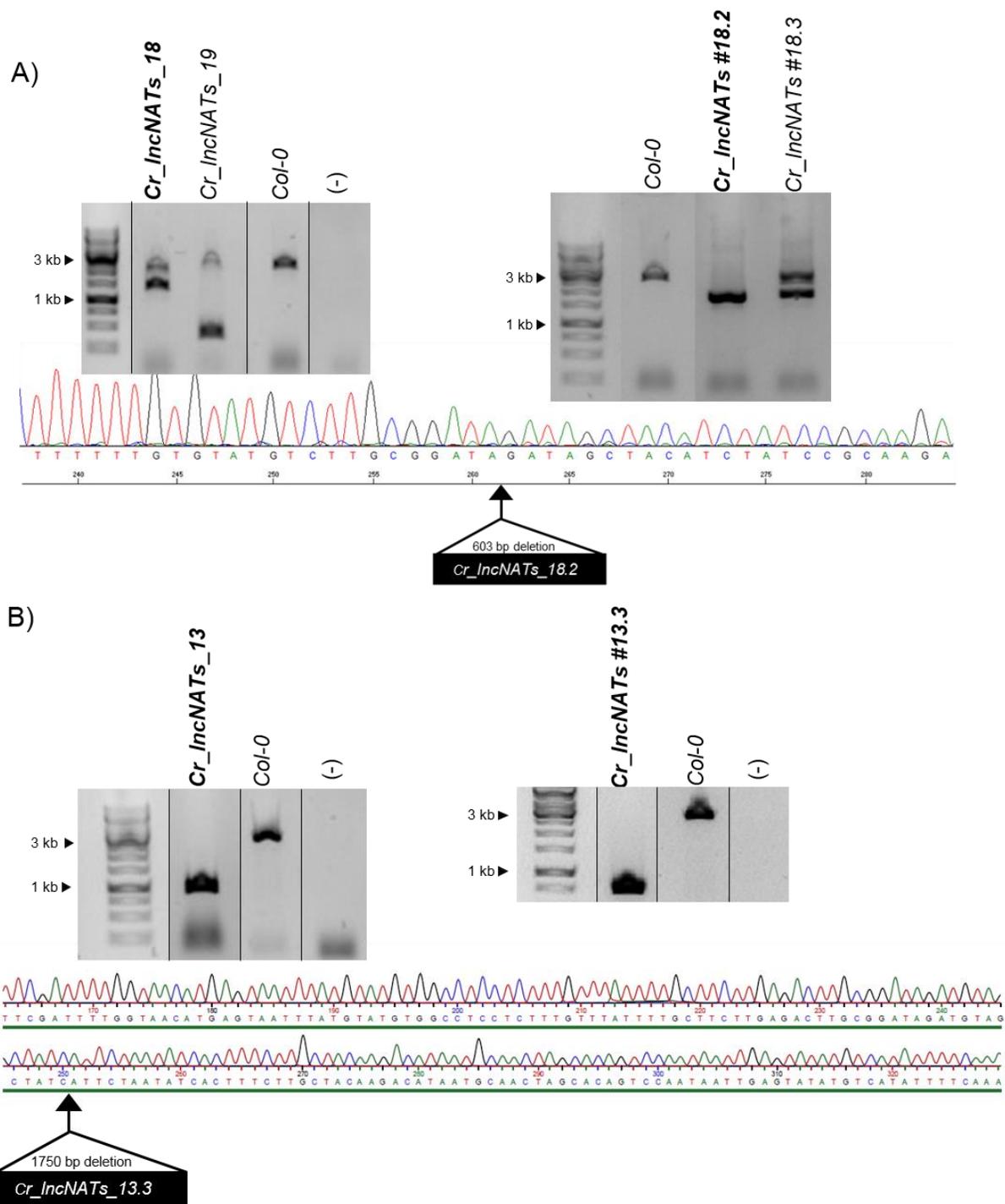
Supplementary Figure 12: **Genotyping of CRISPR/Cas9 knockouts of *lncNATs-UGT73C6*.** A) PCR identification of plants homozygous for partial deletion (603 bp) were identified by PCR at T1 (left) and confirmed at T2 (right) stage. B) PCR showing amplification of *lncNATs-UGT73C6* for plants that were marked as homozygous for full deletion (1750 bp) in T1 (left) and T2 (right). (-) indicate negative control. Chromatograms showing the sequence representing the deletion in *Cr_lncNATs #18.2* and *Cr_lncNATs #13.3* line. Insets indicate the position and length of deletions.

X
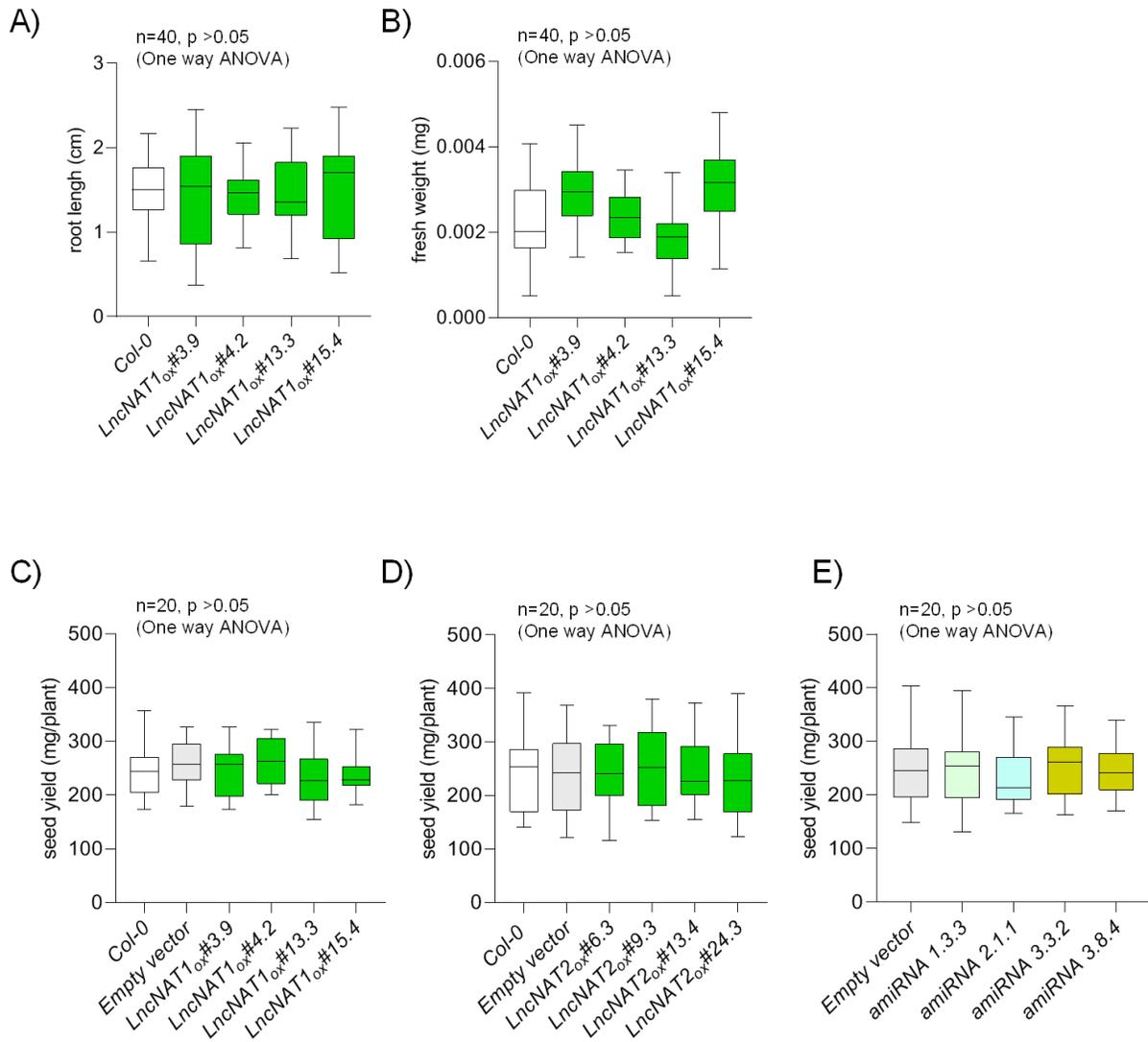
Supplementary Figure 13: **Analysis of phenotypic parameters in lines showing overexpression or downregulation of *lncNATs-UGT73C6*.** Experiments showing effects over primary root length (A) and total fresh weight (B) from 14 DAS seedlings grown over ½ MS agar plates. Representative experiments showing seed yield in lines overexpressing *lncNAT1* (C) or *lncNAT2* (D) and lines in which *lncNATs-UGT73C6* expression is downregulated (E). One representative experiment out of 2 (*lncNAT1_{ox}* lines) and 3 (*lncNAT2_{ox}* and *amiRNA* lines) is shown. Statistical significance, is calculated after comparison to controls transformed with the corresponding empty vectors, details are shown at top of each box plot.

Supplementary Figure 14: **Alignment of *UGT73C5* sequence with *UGT73C6* sequence corresponding to region complementary to purported *lncNAT2* promoter***.* CLUSTAL W sequence alignment showing differences in *UGT73C5* and *UGT73C6* strand at 3' end. Unlike *UGT73C5*, the complementary strand of *UGT73C6* acts as a purported promoter for *lncNAT2*. Nucleotides highlighted in red are different in *UGT73C5* and *UGT73C6*. Yellow colour highlight the sequence compared between *UGT73C5* and *UGT73C6.* Green arrow indicate transcription start site of *lncNAT2* over the complementary strand of *UGT73C6*. Fluorescent downward arrowhead marks the position of nucleotide (1440 bp downstream of *UGT73C6* transcription start site) at which T-DNA is inserted in *ugt73c6<sub>ko</sub>* mutant line that was used in this study*.*

## Tables

Table 1 Sequence complementarity of *lncNATs-UGT73C6* with *UGT73C* family members

| *UGT73C* family member | *LncNAT1 (%)* | *LncNAT2 (%)* | *LncNAT1* vs PSPG Motif (%) |
|---|---|---|---|
| *UGT73C1* | 80 | 77 | 82 |
| *UGT73C2* | 81 | 79 | 79 |
| *UGT73C3* | 83 | 82 | 80 |
| *UGT73C4* | 82 | 80 | 80 |
| ***UGT73C5*** | **91** | **90** | **92** |
| ***UGT73C6*** | **100** | **100** | **100** |
| *UGT73C7* | 69 | 67 | 73 |

Table 2 Rosette area measurements for *lncNAT1* overexpression lines

| Genotype | Empty vector | *LncNAT1$_{ox}$ #3.9* | *LncNAT1$_{ox}$ #4.2* | *LncNAT1$_{ox}$ #13.3* | *LncNAT1$_{ox}$ #15.4* |
|---|---|---|---|---|---|
| | | % increase with respect to empty vector | | | |
| EXP. 1A 25DAS (n-20) | 100 | 9.70* | 7.54 | 9.65 | 5.63 |
| EXP. 2A 25DAS (n=7) | 100 | 20.39* | 14.60* | ND | 12.53 |
| EXP. 1G 25DAS (n=15) | 100 | 18.74* | 21.35* | 9.05 | 10.97* |
| Average (±SD) | | 16.28 (±5.1) | 14.50 (±6.2) | 9.35 (±0.35) | 9.71 (±3.2) |

\* $p < 0.05$, one-way ANOVA, EXP. Experiment, G: greenhouse, A: growth cabinet. Number of plants per lines included in each experiment is indicated with label of each independent experiment. ND: No differences

Table 3 Rosette area measurements for *lncNAT2* overexpression lines

| Genotype | Empty vector | *LncNAT2$_{ox}$ #6.3* | *LncNAT2$_{ox}$ #9.3* | *LncNAT2$_{ox}$ #13.4* | *LncNAT2$_{ox}$ #24.3* |
|---|---|---|---|---|---|
| | | % increase with respect to empty vector | | | |
| EXP. 1G 25DAS (n-20) | 100 | 34.07* | 32.26* | 27.82* | 24.49* |
| EXP. 2G 25DAS (n-20) | 100 | 19.19* | 35.96* | 52.54* | 23.42* |
| EXP. 3G 25DAS (n=15) | 100 | 11.45* | 20.92* | 15.02* | 6.12 |
| Average (±SD) | | 21.57 (±10.2) | 29.74 (±7.0) | 31.79 (±17.0) | 18.01 (±9.2) |

\* $p < 0.05$, one-way ANOVA, EXP. Experiment, G: greenhouse, Number of plants per lines included in each experiment is indicated with label of each independent experiment.

Table 4 Rosette area measurements for *lncNAT2 and mut_lncNAT2* overexpression lines

| Genotype | Empty vector | LncNAT2$_{ox}$ #3.6 | LncNAT2$_{ox}$ #5.1 | LncNAT2$_{ox}$ #6.3 | LncNAT2$_{ox}$ #10.3 | Mut_lnc NAT2$_{ox}$ #9.2 | Mut_lnc NAT2$_{ox}$ #11.2 | Mut_lnc NAT2$_{ox}$ #14.4 | Mut_lnc NAT2$_{ox}$ #16.2 | Mut_lnc NAT2$_{ox}$ #19.1 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | % increase with respect to empty vector | | | | | | | | |
| EXP. 1Vo 25DAS (n-15) | 100 | NA | 23.65* | 30.73* | 30.55* | 15.73* | 13.87* | 12.87* | 35.05* | 14.89* |
| EXP. 2Vo 25DAS (n-20) | 100 | 10.78 | NA | NA | 28.06* | 9.80* | 20.96* | NA | 14.15* | 15.75* |
| * p<0.05, one-way ANOVA, EXP. Experiment, Vo: growth cabinet. Number of plants per lines included in each experiment is indicated with label of each independent experiment. NA: not included in the experiment | | | | | | | | | | |

Table 5 Rosette area experiments for amiRNA downregulation lines

| Genotype | Empty vector (%) | amiRNA 1.3.3 | amiRNA 2.1.1 | amiRNA 3.3.2 | amiRNA 3.3.4 | amiRNA 3.8.4 |
|---|---|---|---|---|---|---|
| | | % decrease with respect to empty vector | | | | |
| EXP. **1G** 25DAS (n-20) | 100 | 26.59* | 10.88* | 32.42* | 20.53* | 8.64 |
| EXP. **2G** 25DAS (n-17) | 100 | 28.2* | 26.62* | 6.49 | NA | 9.04 |
| EXP. **3G** 25DAS (n-17) | 100 | 27.69* | 16.58* | 17.58* | NA | NA |
| **Average (±SD)** | | 27.49 (±0.73) | 18.03 (±7.1) | 18.83 (±11.6) | | |
| EXP. 1A 25DAS (n-17) | 100 | 9.12 | 14.20* | 24.99* | NA | 3.99 |
| * p<0.05, one-way ANOVA, EXP. Experiment, G: greenhouse, A: growth cabinet. Number of plants per lines included in each experiment is indicated with label of each independent experiment. NA: not included in the experiment. Experiments included in analysis are marked in bold | | | | | | |

Table 6 Average number of size of mesophyll cells in *Col-0, lncNAT2$_{ox}$ #10.3* and *mut_lncNAT2$_{ox}$ #16.6*

| | Fixed leaf area (cm$^2$) | Average mesophyll cell size (cm2) | Number of mesophylls/ 2 cm$^2$ |
|---|---|---|---|
| *Col-0* | 2.0 | 2.67 x 10$^{-4}$ $^\$$ | 74923[€] |
| *LncNAT2$_{ox}$ #10.3* | 2.0 | 3.46 x 10$^{-4}$ $^\$$ | 58168[€] |
| *Mut_lncNAT2$_{ox}$ #16.6* | 2.0 | 3.20 x 10$^{-4}$ $^\$$ | 62307[€] |
| $^\$$Average mesophyll cell area = (median cell area at bottom + median cell area at tip)/2<br>[€] Number of mesophyll cells in complete leaf was calculated after diving leaf area by average area of mesophyll cells | | | |

Table 7 Primers Used in the study

| Primer ID | Gene | Primer Sequence | Orientation |
|---|---|---|---|
| DM478 | *β-glucuronidase (GUS)* | 5´ CGTCCTGTAGAAACCCCAAC 3' | Forward |
| DM479 | *β-glucuronidase (GUS)* | 5´ CGGTTTTTCACCGAAGTTCA 3 | Reverse |
| RR433 | *PP2A(AT1G13320)* | 5´ GCTGTAGGACCGGAGCCAACTAG 3´ | Forward |
| RR434 | *PP2A(AT1G13320)* | 5´ CAGGACCAAACTCTTCAGCAAGACGC 3´ | Reverse |
| RR44 | *lncNAT1(At2g36792.1)* | 5´ caccGCTAGTTGCATTATGTCTTGTAG 3´ | Forward |
| RR45 | *lncNAT1(At2g36792.1)* | 5´ GTCTTCTGTTCACTTTCTTTCTTG 3´ | Reverse |
| RR46 | *lncNAT2(At2g36792.2)* | 5´ caccGGGATTCCTCTAGGCCTAGTC 3´ | Forward |
| RR47 | *lncNAT2(At2g36792.2)* | 5´ GGCAAGTAAAAATCTGTTTACG 3 | Reverse |
| RR48 | *Prom::lncNAT1 (At2g36792.1)* | 5´ caccTGCTTCCCCTCTCAGCTTTGTC 3´ | Forward |
| RR49 | *Prom::lncNAT1 (At2g36792.1)* | 5´ caccCATaccgtttgcaacgactcaatga 3´ | Forward |
| RR50 | *Prom::lncNAT2 (At2g36792.2)* | 5´ ACAATTCAAGAATTGAGCATATGTC 3´ | Reverse |
| RR51 | *Prom::lncNAT2 (At2g36792.2)* | 5´ AAAGACCTTTCATCTGGGTCATA 3´ | Reverse |
| RR606 | *LncNAT1 +49--+69* | 5' AGCCTCCTTCTTCCACAGCC 3' | Forward |
| RR607 | *LncNAT1 +108--+129* | 5' CTCTCTTTTGCATCATCACTC 3' | Forward |
| RR608 | *LncNAT1 +183--+205* | 5' GGGGAGAAGAAGAGAAGATAGG 3' | Reverse |
| RR609 | *LncNAT1 +326--+346* | 5' GGAACTCGACTCTTGAGGGG 3´ | Reverse |
| RR445 | *pORF1 lncNAT2* | 5' caccATGGTCCATGCTTTACCAGACC 3' | Forward |
| RR446 | *pORF1 STOP codon removal lncNAT2* | 5' AGAGCTCGAACCTGCGTATGC 3' | Reverse |
| RR447 | *pORF2 lncNAT2* | 5' caccATGTCTTATCCGCTTCTACCATATC 3' | Forward |
| RR448 | *pORF2 STOP codon removal lncNAT2* | 5' CAAGACCTCAAGTTCCGGTGG 3' | Reverse |
| RR449 | *pORF3 lncNAT2* | 5' caccATGAAGTACTCCTTATCAGACTTTAAATTG 3' | Forward |
| RR450 | *pORF3 STOP codon removal lncNAT2* | 5' AGAAGCTGGTCTGCAAGAAGGAC 3' | Reverse |
| RR451 | *pORF4 lncNAT2* | 5' caccATGGGCAAACCAGACTCAATGG 3' | Forward |
| RR452 | *pORF4 lncNAT2* | 5' CAATGCAGCAAGGTTCAAGAATGTC  3' | Reverse |
| RR531 | *pORF1 (2nd)lncNAT2* | 5'GACTATAACACCATAAGCCCTCTTATCCGCTTCTACC 3´ | Forward |
| RR532 | *pORF1 (2nd)lncNAT2* | 5'GGTAGAAGCGGATAAGAGGGCTTATGGTGTTATAGTC 3´ | Reverse |
| RR529 | *pORF2 (3rd)lncNAT2* | 5'CAAGGAAACAGGTCCACCCGTCCATGCTTTACCAG 3´ | Forward |
| RR530 | *pORF2 (3rd)lncNAT2* | 5'CTGGTAAAGCATGGACGGGTGGACCTGTTTCCTTG 3´ | Reverse |
| RR527 | *pORF3 (1st)lncNAT2* | 5'CAGGAAAATAAGGAACACCCAAGTACTCCTTATCAGAC 3´ | Forward |
| RR528 | *pORF3 (1st)lncNAT2* | 5'GTCTGATAAGGAGTACTTGGGTGTTCCTTATTTTCCTG 3´ | Reverse |
| RR533 | *pORF3'(4th)lncNAT2* | 5'CAAAAGCAACCCATGCCCCCGAAGAGGATCTTTGG 3´ | Forward |
| RR534 | *pORF3'(4th)lncNAT2* | 5'CCAAAGATCCTCTTCGGGGGCATGGGTTGCTTTTG 3´ | Reverse |
| RR218 | *amiRNA1 lncNATs-UGT73C6* | 5' TGTATCTGTGCTCTACGTTTGCCTAATGATGATCACATTCGTTATCTATTTTTTAGGCAAACGGAGAGCACAGA 3' | Forward |

| RR219 | *amiRNA1 lncNATs-UGT73C6* | 5' AATGTCTGTGCTCTCCGTTTGCCTAAAAAAATAGATAACGAATGTGATCATCATTAGGCAAACGTAGAGCACAGA 3' | Reverse |
|---|---|---|---|
| RR220 | *amiRNA2 lncNATs-UGT73C6* | 5' TGTATTCGAAAAAAACAACGAACCAATGATGATCACATTCGTTATCTATTTTTTTGGTTCGTTGGTTTTTTCGAA 3' | Forward |
| RR221 | *amiRNA2 lncNATs-UGT73C6* | 5' AATGTTCGAAAAAACCAACGAACCAAAAAAATAGATAACGAATGTGATCATCATTGGTTCGTTGTTTTTTTCGAA 3' | Reverse |
| RR222 | *amiRNA3 lncNATs-UGT73C6* | 5' TGTATAACGTTCTGCGCAAGAACCCATGATGATCACATTCGTTATCTATTTTTTGGGTTCTTGCTCAGAACGTTA 3' | Forward |
| RR223 | *amiRNA3 lncNATs-UGT73C6* | 5' AATGTAACGTTCTGAGCAAGAACCCAAAAAATAGATAACGAATGTGATCATCATGGGTTCTTGCGCAGAACGTTA 3' | Reverse |
| RR236 | *amiRNA1 lncNATs-UGT73C6* | 5´ TAGGCAAACGTAGAGCACAGA 3´ | Anti-*lncNATs-UGT73C6* |
| RR237 | *amiRNA2 lncNATs-UGT73C6* | 5´ TGGTTCGTTGTTTTTTTCGAA 3´ | Anti-*lncNATs-UGT73C6* |
| RR238 | *amiRNA3 lncNATs-UGT73C6* | 5´ GGGTTCTTGCGCAGAACGTTA 3´ | Anti-*lncNATs-UGT73C6* |
| RR173 | *UGT73C5* | 5´ CACCATGGTTTCCGAAACAAC 3´ | Forward |
| RR174 | *UGT73C6* | 5´ TCAATTATTGGGTTCTGC 3´ | Reverse |
| RR283 | *UGT73C6prom::UGT73C6-GUS* | 5' caccCTGGCAGAACCCAATAATTGAGTATACG 3´ | Forward |
| RR284 | *UGT73C6prom::UGT73C6-GUS* | 5' ATTATTGGACTGTGCTAGTTGC 3´ | Reverse |
| RR758 | *DWF4(At3g50660)* | 5´ CATAAAGCTCTTCAGTCACGA 3´ | Forward |
| RR548 | *DWF4(At3g50660)* | 5´ CGTCTGTTCTTTGTTTCCTAA 3´ | Reverse |
| RR551 | *BR6OX2(At5g38970)* | 5´ AGCTTGTTGTGGGAACTCTATCGG 3´ | Forward |
| RR552 | *BR6OX2(At5g38970)* | 5´ CGATGTTGTTTCTTGCTTGGACTC 3´ | Reverse |
| RR557 | *BAS1(At2g26710)* | 5´ CAAATGCTTCTTTGTGCTGAA 3´ | Forward |
| RR558 | *BAS1(At2g26710)* | 5´ aattccctcttgtcggaaaaa 3´ | Reverse |
| RR740 | *ROT3 (At4g36380)* | 5´ CAAATCCTTAAAGGCGAAAGAGAGG 3´ | Forward |
| RR741 | *ROT3 (At4g36380)* | 5´ CTTAGAAGTACGTCCACCACG 3´ | Reverse |
| RR744 | *EXP8 (At2g40610)* | 5´ CAACCATCACCGTCACAGCTA 3´ | Forward |
| RR745 | *EXP8 (At2g40610)* | 5' CTGAAGAGGAGGATTGCACCA 3' | Reverse |
| RR746 | *GRF1 (At2g22840)* | 5´ GGAAAGAAATGGCGGTGCTCG 3´ | Forward |
| RR747 | *GRF1 (At2g22840)* | 5´ CAGCGGCGGCAGCATTAGTA | Reverse |
| RR748 | *GRF2 (At4g37740)* | 5´ GATCAGAAATACTGTGAAAGACACATC 3´ | Forward |
| RR749 | *GRF2 (At4g37740)* | 5´ GTTGTGGTGTAGTAACCGCTTTG 3´ | Reverse |
| RR750 | *GRF3 (At2g36400)* | 5´ TTCGCTGGCCACAAGTATTGC 3´ | Forward |
| RR751 | *GRF3 (At2g36400)* | 5´ TGTTGCTGTTGTAGTGGTGGCT 3´ | Reverse |
| RR752 | *GRF4 (At3g52910)* | 5´ GCGGGCCACAAGTATTGTGAC 3´ | Forward |
| RR753 | *GRF4 (At3g52910)* | 5´ GGAGAAGAAGTGGTTGTTGTTTGG 3´ | Reverse |
| RR754 | *GRF5 (At3g13960)* | 5´ CTCTTCATCATGCTTCCGCTTT 3´ | Forward |
| RR755 | *GRF5 (At3g13960)* | 5´ TTGCTAACGGTTGTTGGTGATG 3´ | Reverse |
| RR756 | *GRF6 (At2g06200)* | 5´ TCCTCAAGAAAGCCTCCTCCTA 3´ | Forward |
| RR757 | *GRF6 (At2g06200)* | 5´ ATCTTCCATTGCTGAGCCAGAG 3´ | Reverse |
| RR758 | *GRF7 (At5g53660)* | 5´ TGCGCGAAAGAAGTCGTCTCTA 3´ | Forward |
| RR759 | *GRF7 (At5g53660)* | 5´ CACCATTGTTGTTAGGGCGAGA 3´ | Reverse |
| RR760 | *GRF8 (At4g24150)* | 5´ GCATGTGGAATCATCTCACCAATC 3´ | Forward |
| RR761 | *GRF8 (At4g24150)* | 5´ GTGATCATATGGAGAGGAGGC 3´ | Reverse |
| RR762 | *GRF9 (At2g45480)* | 5´ AGAAGCTGCAAAATTCTCAAGCAAC 3´ | Forward |
| RR763 | *GRF9 (At2g45480)* | 5´ GCACCTTCCTGGTTCATTGTC 3´ | Reverse |
| *RR882* | *AN3 (At5g28640)* | 5´ CTGATCATATCCAACAGTACTTGG 3´ | Forward |
| *RR883* | *AN3 (At5g28640)* | 5´ GCCTTGCTTGATTCTCGGCG 3´ | Reverse |

| RR289 | *UGT73C1* | 5´ TACTTCCCCATTCCTAATTTCCCTGAC 3´ | Forward |
|---|---|---|---|
| RR290 | *UGT73C1* | 5´ CTGGCTCGAGCTCTTCAAACGTGTTA 3´ | Reverse |
| RR291 | *UGT73C2* | 5´ CTACACCGAAACCACAATATCTTACATG 3´ | Forward |
| RR292 | *UGT73C2* | 5´ GGACGTGTCATCAGCATCCACCTG 3´ | Reverse |
| RR293 | *UGT73C3* | 5´ CGAGCGATCGAGTCTGGCTTGG 3´ | Forward |
| RR294 | *UGT73C3* | 5´ AAGGTACCATCAACTCCGTTGAGTCT 3´ | Reverse |
| RR295 | *UGT73C4* | 5´ CGCAGAAACCTCGAGATCTTGAAGAAC 3´ | Forward |
| RR296 | *UGT73C4* | 5´ AAGAACGCTTTCCAATCTCCACTTGCAG 3´ | Reverse |
| RR353 | *UGT73C5* | 5´ GGGGATAACTGCTGGTCTACCG 3´ | Forward |
| RR354 | *UGT73C5* | 5´ CCTCTTCTCCCCATTTCATAGGCTG 3´ | Reverse |
| RR287 | *UGT73C6 (At2g36790)* | 5´ GAAGCTGGTCTGCAAGAAGGACAAGAA 3´ | Forward |
| RR288 | *UGT73C6 (At2g36790)* | 5´ GATTAGACAGCTTGGTCGCGGGC 3´ | Reverse |
| RR297 | *UGT73C7* | 5´ CTCTAGGCTCTTGTCCCAGCGCCAA 3´ | Forward |
| RR298 | *UGT73C7* | 5´ CGCACCCTTCTGGCAAACCCGTTTG 3´ | Reverse |
| RR357 | *LncNATs-UGT73C6 (At2g36792)* | 5´ GGCAAACGTAGAGCACAGATCCCG 3´ | Forward |
| RR358 | *LncNATs-UGT73C6 (At2g36792)* | 5´ GTAGGAGTAGACAAAGCAGAGAGGGG 3´ | Reverse |
| DM195 | *PP2A(AT1G13320)* | 5' AGCCAACTAGGACGGATCTGGT 3' | Forward |
| DM196 | *PP2A(AT1G13320)* | 5' GCTATCCGAACTTCTGCCTCATTA 3' | Reverse |
| RR574 | *At3g45970 (EXPANSIN L1)* | 5´ CAAGTCGGTTCATCGCCAAATTGGG 3´ | Forward |
| RR575 | *At3g45970 (EXPANSIN L1)* | 5´ GTATCCACCGGTTACTACGAACCTG 3´ | Reverse |
| RR576 | *AT3G13920 (EIF4A1)* | 5´ GGCGTAAGGTTGATTGGCTCAC 3´ | Forward |
| RR577 | *AT3G13920 (EIF4A1)* | 5´ GATGAGAACACGGGAGGAACCAG 3´ | Reverse |
| RR242 | *Upstream of the pB7WG2 attR1 site* | 5´ TCGACCTGCAGGCGGCCGC 3´ | Forward |

## Acknowledgements

First of all, I would like to express my sincere gratitude to my PhD supervisor Dr. Selma Gago Zachert for her continuous and all-round support from the very first day of my arrival in Germany. Your, patience, perseverance, immense motivation, knowledge, amazing capabilities of supervision, continuous thrust for perfection and particularly your unpretentious nature truly helped me to get an essence of '*what science is and how it should be pursued without any compromise'*. Surely, your paramount dedication to science and curiosity driven logical attitude helped me a lot in my holistic evolution while working on a PhD project that was 'essentially' started from the very scratch. It's been a privilege and indeed a life-time experience that I did my PhD under your supervision. Thank you so much from the bottom of my heart.

Further, I am greatly thankful to Prof. Dr. Steffen Abel for hosting me as a PhD student and for reviewing my thesis. Also, your continuous encouragements and support especially during presentations has been very much helpful in the building up of my confidence level. I am also wholeheartedly thankful to Prof. Dr. Sven-Erik Behrens for reviewing my thesis and not only for being my mentor but also for the support as coordinator of the Erasmus Mundus Action-2 BRAVE program of the European Union. In addition, I am obliged to you for facilitating and optimistically supporting the investigations during the entire period of my PhD project.

My PhD experience would have not been in the present form without the friendly company and cooperative aptitude of lab colleagues. In this regard, thanks to Ammar, Susane and Katja for contributing to the completion of my PhD project and for maintaining the cooperative atmosphere in laboratory. Particularly, I am thankful to Ammar not only for helping with German language but also for being nice, innovative, helpful, and giving me time to help solve the problems during the difficult situations within and outside PhD domain. Moreover, I am grateful to previous students in lab namely, Tebbe de Vries and Michael Fritz for performing preliminary experiments as part of their dissertations. Many thanks to Michel Heidecker for the nice company, discussions and for your keen interest in the extensive efforts to 'disprove' the target mimicry hypothesis. I also acknowledge Saskia for RACE analysis and Emilia for attempting to contribute for RNA-protein interaction assays. Thanks to all MSV and IPB colleagues for the scientific environment in the department and institute, respectively. I am greatly thankful to gardeners for taking care of plants in the greenhouse. My heartfelt appreciation goes to all those who directly or indirectly contributed to the progress of my research project.

I would also like to acknowledge BRAVE coordinator Prof. Andreas Voloudakis (Agricultural University, Athens, Greece) for his constant support and funding during for the first three years (2014-2017). Furthermore, thanks to IBP and GRK-1591 for financial support during 2017-2019 period.

A special mention to Dr. Rahul Pal for entrust and the overwhelming support while I was in India. Also, I am grateful to Dr. Apurba Kumar Sau for being my supervisor and developing my interest in enzyme kinetics during my stay at NII, New Delhi.

Besides, my experience of living in Halle and Germany as a whole would have not been enjoyable and memorable without the company of *BRAVE* colleagues and *Halle Indians*. Special

thanks to my friend Dr. Mayank for making my stay in Germany memorable and for being a charming companion and fellow-explorer. Also, I greatly appreciate company of *fellas* viz: Dr. Prodyut (MLU, Pharmacy), Dr. Ajeet (TUM, Munich) and Dr. Tamil (including his family).

I am especially thankful to Dr. Neeraj and Dr. Gupta for being my confidantes and friends in the process of co-evolution of our personalities and for imparting the motivation to pursue science as a career since the BHU and JNU days. Thank you doctors. Also, I am vastly thankful to Sanskrit Professor Mohar Singh Meena (elder brother and cousin) for teaching me a lot and taking care of me since my childhood. Your advice that 'साहित्य में बहुत ग्यान छुपा हुआ है, शिव' (literature is the real source of knowledge, Shiv) always fascinates me.

Last but not least, I am truly beholden to my family. It's because of your patience, support, care, love and **'*hopes*'** that I could accomplish my thesis. Thanks to my siblings Ajay and Deepa for being there with me. I owe a deep sense of indebtedness to my Papa (Rajendra Singh) and Mammi (Jal Devi) for being wonderful parents, believing in me and acting as ladder all the time. Likewise, I am immeasurably indebted to Priyanka and her family. I have no words to express my love and feelings for your sacrifice and company. I could not have imagined a partner better than you. Thank you so much for holding the grounds all the time and providing the paddles while I am sailing the boat through the journey of academic dreams and life.
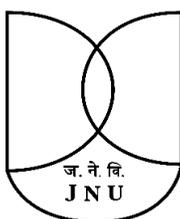
Finally, I am thankful to the Almighty God for the endless blessings.

# Curriculum vitae

Name: **Shiv Kumar Meena**
Nationality: Indian

## Education

| | |
|---|---|
| 2014 | **Doctor rerum naturalium** (**Dr. rer. nat.**), Leibniz Institute for Plant Biochemistry, Halle (Saale), Germany<br>Thesis: Analysis of *UGT73C6* natural *cis*-antisense long non-coding RNAs in *Arabidopsis thaliana* |
| 2010 | **Master of Life Sciences**, Jawaharlal Nehru University, New Delhi, India, Thesis: Evaluation of radio-modulatory potential of naringenin, |
| 2008 | **Bachelor of Sciences**, Botany (Honors), Banaras Hindu University, Varanasi, India, Thesis: Evaluation of geraniol as a botanical preservative against fungal and insect infestation |

## Work Experience

| | |
|---|---|
| July 2010 – August 2013 | **Junior Research Fellow**, National Institute of Immunology (NII), New Delhi, India, Project: Structural, functional and kinetic studies of *Helicobacter pylori* arginase, Supervisor : Dr. Apurba Kumar Sau |

## Awards/Fellowships

| | |
|---|---|
| 2018- 2019 | Graduiertenkolleg (GRK-1591) (Deutsche Forschungsgemeinschaft) |
| 2014 – 2017 | Erasmus Mundus PhD Fellowship |
| 2010 – 2013 | NII, Delhi, Research Fellowship |
| 2010 | Qualified all India level "Graduate Aptitude Test in Engineering (GATE)" conducted by Indian Institute of Technology, Delhi, India, Stream: Life Sciences |
| 2009 | Junior Research Fellowship and lectureship (JRF) cum National Eligibility Test (NET) by Council of Scientific and Industrial Research, India |
| 2008 | Merit-Cum-Means M.Sc. Scholarship in Life Sciences |

## Publication(s) in peer reviewed journals

Srivastava A\*, **Meena SK\*,** Alam M, Nayeem SM, Deep S, Sau AK (2013) Structural and functional insights into the regulation of *Helicobacter pylori* arginase activity by an evolutionary non-conserved motif. *Biochemistry* 2013 52 (3), 508-519. (\*-Co first author)

## Unpublished work

**Meena SK**, Jaber A, Höpfner S, de Vries T, Behrens SE, Abel S and Gago-Zachert S, A nonconserved antisense long noncoding RNA modulates leaf area in *Arabidopsis thaliana* (to be submitted)

## Presentations in Congresses and Meetings

**Meena SK\*,** Jaber A, de Vries T, Höpfner S, Behrens SE, Abel S, Gago-Zachert S (2019) *Characterization of Arabidopsis thaliana UGT73C6 natural cis-antisense long non-coding RNAs and analysis of their role in leaf size modulation.* RNA 2019, Krakow, Poland

**Meena SK\*,** de Vries T, Höpfner S, Jaber A, Behrens SE, Abel S, Gago-Zachert S (2017) *Alterations in the expression levels of two antisense long non-coding RNAs overlapping a UGT73C6 gene of Arabidopsis thaliana results in changes in leaf area.* RNA 2017, Prague, Czech Republic

**Halle (Saale), August 2020**                                                                                           **Shiv Kumar Meena**

XX

# Statuary declaration

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 an der Saale                                                                          Shiv Kumar Meena