

# **Transcription elongation factors and the chromatin status affect alternative splicing in *Arabidopsis Thaliana***



**DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES  
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN  
DER UNIVERSITÄT REGENSBURG**

**vorgelegt von  
Silvia Esposito  
aus  
Pistoia, Italy  
im Jahr 2020**

Das Promotionsgesuch wurde eingereicht am:

10.07.2020

Die Arbeit wurde angeleitet von:

Prof. Dr. Klaus D Grasser

Unterschrift:

Silvia Esposito





# Contents

<b>ABBREVIATIONS</b> .....	<b>IV</b>
<b>SUMMARY</b> .....	<b>1</b>
<b>1. INTRODUCTION</b> .....	<b>3</b>
1.1 TRANSCRIPTION BY RNA POLYMERASE II .....	3
1.2 TRANSCRIPTION ELONGATION .....	4
1.3 CO-TRANSCRIPTIONAL SPLICING OF MRNAs.....	6
1.4 MRNA SPLICING IN PLANTS .....	8
1.5 ALTERNATIVE SPLICING.....	9
1.6 ALTERNATIVE SPLICING: HUMANS VS. PLANTS .....	10
1.7 COUPLING BETWEEN ALTERNATIVE SPLICING AND TRANSCRIPTION .....	10
1.8 KINETIC COUPLING VS. RECRUITMENT COUPLING .....	11
1.9 THE CHROMATIN STRUCTURE AS A REGULATOR OF AS.....	14
1.10 THE ROLE OF H3K36ME3, MRG PROTEINS AND PTB IN THE REGULATION OF ALTERNATIVE SPLICING .....	14
<b>AIM OF THE PROJECT</b> .....	<b>17</b>
<b>2. RESULTS: THE EFFECT OF TRANSCRIPTION ELONGATION FACTORS ON ALTERNATIVE SPLICING IN ARABIDOPSIS THALIANA</b> .....	<b>19</b>
2.1 THE <i>TEF</i> MUTANTS <i>SPT4R3</i> , <i>TFIIS</i> , <i>IWS1</i> AND <i>TFIIS/IWS1</i> SHOW ALTERNATIVE SPLICING DEFECTS .....	19
2.2 RNASEQ DATA ANALYSIS REVEALS DIFFERENTIALLY ALTERNATIVELY SPLICED GENES IN <i>TEF</i> MUTANTS .....	19
2.3 GENE ONTOLOGY ANALYSIS OF GENES AFFECTED IN AS .....	23
2.4 EXPERIMENTAL VALIDATION OF DIFFERENTIALLY ALTERNATIVELY SPLICED GENES .....	25
2.5 GENOME-WIDE ANALYSIS OF RNAPII OCCUPANCY: FROM RNASEQ TO CHIPSEQ.....	29
2.6 QUALITY CONTROL OF CHIP EXPERIMENTS.....	30
2.7 RNAPII OCCUPANCY IN <i>TEF</i> MUTANTS .....	31
2.8 THE PHOSPHORYLATION STATUS OF RNAPII IS CORRELATED WITH PAUSING AT INTRON/EXON BOUNDARIES .....	34
2.9 VALIDATION OF RNAPII CHIPSEQ RESULTS BY CHIP-QPCR.....	37
<b>3. RESULTS: THE CHROMATIN STATUS AND ALTERNATIVE SPLICING</b> .....	<b>41</b>
3.1 THE <i>ARABIDOPSIS</i> H3K36ME3 – MRG ADAPTOR COMPLEX.....	41
3.2 SPLICING FACTORS AND THE METHYLTRANSFERASE SDG7 CO-PURIFY WITH MRG PROTEINS ....	41
3.3 MRG PROTEINS DIRECTLY INTERACT WITH THE METHYLTRANSFERASE SDG7 .....	47
3.4 MOLECULAR CHARACTERIZATION OF <i>MRG</i> AND <i>IWS1</i> T-DNA INSERTION LINES .....	50
3.5 PHENOTYPIC ANALYSIS OF <i>MRG</i> AND <i>IWS1</i> MUTANT LINES.....	52
3.6 MRG PROTEINS DO NOT INFLUENCE AS PATTERN OF SELECTED AS GENES.....	54
3.7 GLOBAL LEVELS OF H3K36ME3 ARE REDUCED IN <i>IWS1</i> MUTANT PLANTS.....	56
3.8 GLOBAL H3K36ME3 DENSITY IS REDUCED IN <i>TEF</i> MUTANT PLANTS .....	57
3.9 VALIDATION OF H3K36ME3 CHIPSEQ RESULTS BY CHIP-QPCR .....	60
3.10 RECRUITMENT OF MRG2 IN <i>IWS1</i> MUTANT PLANTS .....	62
<b>4. DISCUSSION: THE INFLUENCE OF TRANSCRIPTION ELONGATION FACTORS ON ALTERNATIVE SPLICING</b> .....	<b>65</b>
4.1 DIFFERENT TEFs AFFECT ALTERNATIVE SPLICING OF DIFFERENT TARGETS .....	66
4.2 THE ABSENCE OF TEFs AFFECTS ALTERNATIVE SPLICING OF SPLICING REGULATORS.....	67
4.3 INTRON RETENTION IS THE MOST PREVALENT AS EVENT ALTERED IN THE <i>TEF</i> MUTANTS .....	68

4.4 DIFFERENT TEFs AFFECT ALTERNATIVE SPLICING OF DIFFERENT TARGETS WITH SIMILAR BIOLOGICAL FUNCTIONS .....	69
4.5 THE ABSENCE OF TEFs INFLUENCES RNAPII OCCUPANCY ALONG GENES .....	69
4.6 RNAPII PHOSPHORYLATED AT SER5 ACCUMULATES AT INTRON/EXON BOUNDARIES .....	71
<b>5. DISCUSSION: THE H3K36ME3 – MRG ADAPTOR COMPLEX .....</b>	<b>71</b>
5.1 MRG PROTEINS CO-PURIFY WITH SPLICING FACTORS.....	72
5.2 MRGs CO-PURIFY WITH THE METHYLTRANSFERASE SDG7 .....	73
5.3 IWS1 AND MRG PROTEINS MAY ACT IN THE SAME PATHWAY .....	74
5.4 IWS1 AFFECTS GLOBAL H3K36ME3 LEVELS .....	76
5.5 THE LOSS OF IWS1 MAY RESULT IN LESS BINDING OF MRG2 TO THE CHROMATIN .....	77
<b>6 MATERIALS.....</b>	<b>79</b>
6.1 INSTRUMENTS .....	79
6.2 CHEMICALS, SOLUTIONS AND ENZYMES .....	79
6.3 OLIGONUCLEOTIDES .....	80
6.4 PLASMIDS .....	84
6.5 T-DNA AND RNAI LINES .....	85
6.6 DATABASES, ONLINE TOOLS AND SOFTWARE .....	85
<b>7 METHODS.....</b>	<b>87</b>
7.1 NUCLEIC ACID BASE METHOD .....	87
7.1.1 Genomic DNA isolation from Arabidopsis leaves .....	87
7.1.2 RNA isolation from Arabidopsis leaves (for alternative splicing experiment) .....	87
7.1.3 DNase-digest.....	87
7.1.4 Reverse transcription (cDNA synthesis).....	88
7.1.5 Polymerase chain reaction .....	88
7.1.6 Real time quantitative PCR (qRT-PCR) .....	89
7.1.7 Reverse Transcriptase PCR (RT-PCR) for alternative splicing analysis .....	89
7.1.8 Calculation of Splicing isoforms.....	89
7.1.9 Alternative splicing PCR isoforms sequencing .....	90
7.1.10 Agarose gel electrophoresis .....	90
7.1.11 DNA extraction from Agarose gel and PCR clean up.....	90
7.1.12 Restriction digest and dephosphorylation.....	90
7.1.13 Ligation .....	90
7.1.14 Isolation of plasmid DNA from E. coli .....	90
7.1.15 Sequencing.....	91
7.2 RNASEQ .....	91
7.3 CHIP SEQ.....	92
7.3.1 Material preparation.....	92
7.3.2 DNA quality control for ChIP seq.....	93
7.3.4 Library preparation and sequencing .....	94
7.4 CHIP SEQ DATA ANALYSIS .....	94
7.5 PROTEIN BASED METHODS .....	94
7.5.1 Protein extraction from Arabidopsis seedlings .....	94
7.5.2 SDS-PAGE .....	95
7.5.3 Coomassie Brilliant Blue (CBB) staining .....	95
7.5.4 Western Blot .....	95
7.4.5 SG-tagged Protein purification .....	96
7.4.6 Trypsin in gel-digestion and purified proteins .....	96
6.6 MICROBIAL WORK .....	97
7.6.1 Cultivation of bacteria .....	97
7.6.2 Preparation of chemically competent cells .....	97
7.6.3 Transformation of chemically competent E. coli cells.....	98
7.6.4 Transformation of competent A. tumefaciens cells.....	98
7.6.5 Production of chemically competent yeast cells .....	98

7.6.6 Transformation of yeast competent cells .....	98
7.6.7 Yeast-2-Hybrid assay .....	99
7.7 PLANT WORK .....	99
7.7.1 Cultivation of Arabidopsis plants .....	99
7.7.2 Transformation of Arabidopsis plants by floral dipping .....	99
7.7.3 Crossing of Arabidopsis plants .....	100
7.7.4 Soil-based phenotyping of Arabidopsis plants .....	100
7.7.5 Arabidopsis plants grown on plates for alternative splicing analysis .....	100
7.7.6 Cultivation of Arabidopsis PSB-D cells .....	100
7.7.7 Transformation of Arabidopsis PSB-D cells .....	101
7.7.8 Infiltration of Nicotiana benthamiana leaves .....	102
7.8 MICROSCOPY .....	102
7.8.1 Confocal Laser Scanning Microscopy (CLSM) .....	102
7.8.2 Förster resonance energy transfer (FRET) .....	102
<b>8. SUPPLEMENTS .....</b>	<b>103</b>
<b>LIST OF FIGURES .....</b>	<b>147</b>
<b>LIST OF TABLES .....</b>	<b>149</b>
<b>PUBLICATIONS .....</b>	<b>151</b>
<b>BIBLIOGRAPHY .....</b>	<b>153</b>
<b>ACKNOWLEDGMENTS .....</b>	<b>169</b>

## Abbreviations

(v/v)	Volume per volume
(w/v)	Weight per volume
3' SS	3' splice site
5' SS	5' splice site
<i>A. Thaliana</i>	<i>Arabidopsis Thaliana</i>
<i>A. Thumefaciens</i>	<i>Agrobacterium Thumefaciens</i>
aa	Amino acid
ACT2	Actin 2
ANOVA	Analysis of variance
AP-MS	Affinity purification coupled to mass spectrometry
AS	Alternative splicing
as	After sonication
bp	base pair
bs	Before sonication
CBB	Coomassie brilliant blue
CBC	Cap binding complex
CBP20	Cap binding protein 20
CBP80	Cap binding protein 80
cDNA	complementary DNA
CDS	Coding sequence
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation sequencing
CLSM	Confocal laser scanning microscopy
Col-0	Columbia-0
CS	constitutive splicing
CTD	C-terminal domain
DAS	Days after stratification
DDO	double drop out medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DOG1	delay of germination 1
DSIF	DRB-sensitivity inducing factor
DTT	Di-thiotreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylene diamine tetraacetic acid
eGFP	enhanced green fluorescent protein
EI	Exitron
ELF 7	EARLY FLOWERING 7
ES	Exon skipping
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FACT	Facilitates chromatin transcription
FLC	flowering locus C
FRET	Förster resonance energy transfer



<b>FT</b>	flowering locus T
<b>GAP</b>	Glycerinaldehyde-3-phosphate dehydrogenase
<b>gDNA</b>	Genomic DNA
<b>GFP</b>	Green fluorescence protein
<b>GK</b>	Gabi-kat
<b>GO</b>	Gene ontology
<b>GRO-seq</b>	Global run-on sequencing
<b>GTF</b>	General transcription factor
<b>HAT</b>	Histone acetyltransferase
<b>HMT</b>	Histone methyltransferase
<b>hnRNP</b>	Heterogenous nuclear ribonucleoprotein
<b>HPR1</b>	Hyper recombinant 1
<b>HYG</b>	Hygromycin B
<b>IgG</b>	Immunoglobuline G
<b>IR</b>	intron retention
<b>ISE</b>	Intronic splicing enhancer
<b>ISS</b>	Intronic splicing silencer
<b>IWS1</b>	Interact with SPT6
<b>Kan</b>	Kanamycin
<b>kDa</b>	kilo Dalton
<b>LB</b>	Left border or Luria Bertani
<b>LD</b>	Long day
<b>miRNA</b>	MicroRNA
<b>MOS11</b>	Modifier of SNC1, 11
<b>MRG15</b>	MORF-related gene on chromosome 15
<b>mRNA</b>	Messenger RNA
<b>mRNP</b>	Messenger ribonucleoprotein
<b>MS</b>	Murashige-Skoog or mass spectrometry
<b>MXE</b>	Mutually exclusive exon
<b>N. benthamiana</b>	Nicotiana benthamiana
<b>NASC</b>	Nottingham Arabidopsis stock centre
<b>NELF</b>	Negative elongaation factor
<b>NET-seq</b>	native elongating sequencing
<b>NLS</b>	Nuclear localization signal
<b>NoAb</b>	No Antibody
<b>NPC</b>	Nuclear pore complex
<b>NUPs</b>	Nucleoporin protein
<b>NXT1</b>	NTF2-related export protein 1
<b>o/n</b>	Over night
<b>OD600</b>	Optical density measured at 600 nm
<b>ORF</b>	Open reading frame
<b>ORI</b>	Origin of replication
<b>P-TEFb</b>	positive transcription elongation factor
<b>PAF1-C</b>	Polymerase-associated factor 1 complex
<b>PAGE</b>	Polyacrylamide gel electrophoresis
<b>PCR</b>	Polymerase chain reaction
<b>PIC</b>	pre-initiation complex
<b>PMSF</b>	Phenylmethysulphonyl fluoride
<b>polyA</b>	Polyadenylation site
<b>PTB2</b>	polypyrimidine tract-binding protein

<b>PTMs</b>	post-translational modifications
<b>QDO</b>	quadruple drop out medium
<b>qRT-PCR</b>	Quantitative reverse transcription-polymerase chain reaction
<b>RB</b>	Right border
<b>RBD</b>	RNA binding domain
<b>RBP1</b>	RNAP II largest subunit
<b>RFP</b>	Red fluorescent protein
<b>RNA</b>	Ribonucleic acid
<b>RNA-seq</b>	Ribonucleic acid sequencing
<b>RNAP I</b>	RNA polymerase I
<b>RNAPII</b>	RNA polymerase II
<b>RNAPII</b>	RNA polymerase II
<b>RNAPIII</b>	RNA polymerase III
<b>ROI</b>	Region of interest
<b>Rpm</b>	Rounds per minute
<b>RRMs</b>	RNA recognition motives
<b>RT</b>	Room temperature
<b>RT-PCR</b>	Reverse transcription-polymerase chain reaction
<b><i>S. cerevisiae</i></b>	<i>Saccharomyces cerevisiae</i>
<b>SAIL</b>	Syngenta Arabidopsis Insertion Library
<b>SBP</b>	Streptavidin binding peptide
<b>SDG7</b>	SET domain group 7
<b>SDG8</b>	SET domain group 8
<b>SDS-PAGE</b>	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
<b>Ser2</b>	Serine 2 phosphorylated CTD
<b>Ser5</b>	Serine 5 phosphorylated CTD
<b>SETD2</b>	SET domain methyltransferase
<b>SG Tag</b>	consisting of SBP and 2x Protein G domains
<b>siRNA</b>	small interfering RNA
<b>snRNA</b>	Small nuclear RNA
<b>SPT16</b>	Suppressor of Ty 16, FACT subunit
<b>SPT4</b>	suppressor of ty 4/5
<b>SPT5</b>	Suppressor of Ty 5
<b>SR</b>	Serine/Arginine-rich proteins
<b>T-DNA</b>	Transfer DNA
<b>tasiRNA</b>	Trans-acting siRNA
<b>TEC</b>	Transcription elongation complex
<b>TEF</b>	Transcription elongation factor
<b>TES</b>	Transcription termination site
<b>TEV</b>	Tobacco Etch Virus protease
<b>TSS</b>	Transcription start site
<b>UAP56</b>	U2AF65 associated protein
<b>UTR</b>	Untranslated region
<b>WT</b>	Wild type
<b>Y2H</b>	Yeast-two-Hybrid

## Summary

Alternative splicing (AS) can be regulated in various ways. In this study it could be demonstrated that the loss of transcription elongation factors (TEFs) results in altered splicing pattern of a variety of genes. Different TEFs were shown to affect splicing of distinct targets with similar biological functions. Among the differentially alternatively spliced genes, intron retention was the most prevalent altered AS event highlighting a main difference in AS between plant and mammalian systems. Two models have been proposed to explain the coupling between transcription and alternative splicing: kinetic coupling and recruitment coupling. According to the kinetic model, the RNAPII elongation rate affects the capacity of the spliceosome to recognize weaker splice sites what affects the splicing decision. In the recruitment model, histone modifications, chromatin status and RNAPII are key players in recruiting splicing factors to the transcription elongation complex (TEC) by an interaction with the C-terminal domain (CTD) of RNAPII. Chromatin Immunoprecipitation sequencing (ChIPseq) revealed that the TEFs Suppressor of Ty 4 (SPT4) and Interacts with SPT6 (IWS1) influence RNAPII occupancy along genes providing an explanation for the altered splicing pattern in the *tef* mutants based on the kinetic model. Additionally, ChIPseq disclosed that SPT4 and IWS1 affect the phosphorylation status of RNAPII in different ways what according to the recruitment model, may explain the fact that both TEFs influence AS of distinct subsets of genes.

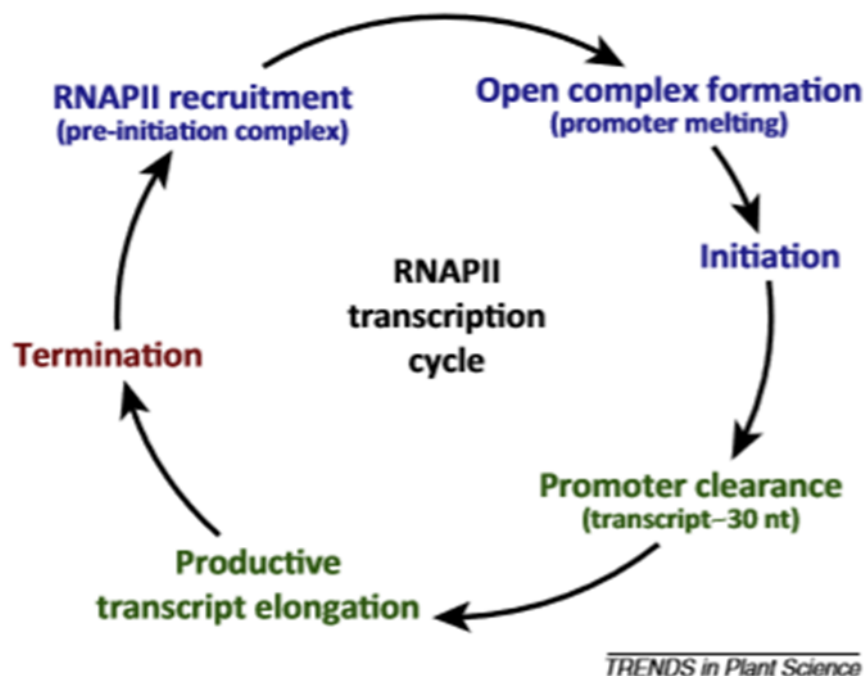
Besides the RNAPII processivity also chromatin adapter complexes can influence splicing decisions. The chromatin reader proteins MORF-related gene 1 and 2 (MRG1 and MRG2) co-purified with various splicing related proteins but not with the polypyrimidine tract-binding protein (PTB2), indicating that in *Arabidopsis* an H3K36me3 adaptor complex is involved in the regulation of splicing but in a different way compared to the situation in mammals. The methyltransferase SET DOMAIN GROUP 7 (SDG7) was identified as a putative interactor of MRG proteins and a direct protein-protein interaction between MRG1, MRG2 and SDG7 could be confirmed by yeast-two hybrid and FRET experiments, demonstrating that SDG7 may be involved in the regulation of an MRG dependent chromatin adapter complex. Phenotyping of plants lacking the TEF IWS1 and both MRG factors suggests that IWS1 and MRGs function in the same pathway what points to a similar regulatory mechanism of the H3K36me3 – MRG chromatin adapter complex in mammals and plants. Global reduced H3K36me3 levels in *iws1* mutant plants observed by Western blotting and ChIPseq further highlighted that also in plants IWS1 is a key regulator of H3K36me3 deposition. ChIP experiments immunoprecipitating MRG2 in *iws1* mutant background further showed a correlation between H3K36me3 histone mark deposition and MRG2 recruitment to chromatin, indicating that in *Arabidopsis* an IWS1 regulated H3K36me3 – MRG chromatin adapter complex is involved in the regulation of AS.



# 1. Introduction

## 1.1 Transcription by RNA polymerase II

Transcription is the most regulated part of gene expression and the RNA polymerases are in the centre of this process. In eukaryotes, protein-coding genes are transcribed by RNA polymerase II (RNAPII) in a three-step transcription cycle consisting of transcription initiation, transcription elongation and transcription termination (Figure 1). Before transcription is initiated, RNAPII together with general transcription factors (GTFs) form a pre-initiation complex (PIC) which binds to the promoter region of a gene. After melting and separation of the two DNA strands, RNAPII starts the synthesis of RNA from the DNA template. During the transition from transcription initiation to elongation, the RNAPII complex has to undergo different structural and functional maturation steps. At first, it has to break its contacts with the promoter site (promoter clearance), before approximately 30 nucleotides downstream of the transcription start-site (TSS) the complex reaches complete stability (Saunders et al., 2006). Before entering productive transcription elongation, RNAPII may be additionally subjected to promoter-proximal pausing, a mechanism that is widely described in metazoans but less well-marked in plants (Jonkers and Lis, 2015; Saunders et al., 2006). Several proteins are involved at this stage and they can either promote or stop RNAPII elongation (Kwak and Lis, 2013). After productive elongation, transcription is terminated by mRNA cleavage and polyadenylation. After transcription termination, the polymerase dissociates from the DNA and can be recycled for a next round of transcription (Van Lijsebettens and Grasser, 2014; Saunders et al., 2006).

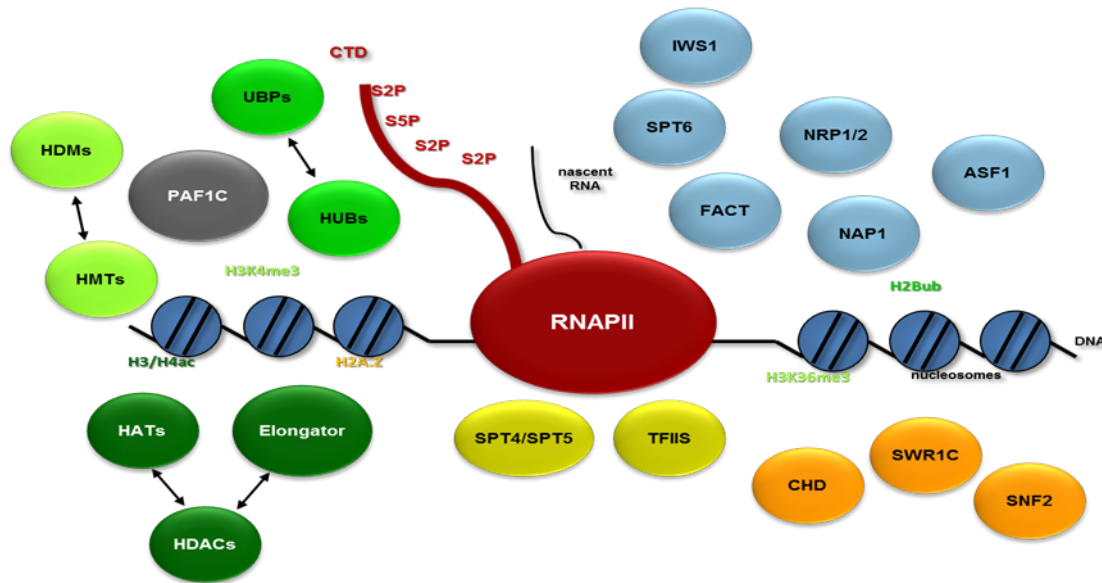


**Figure 1. RNAPII transcription cycle.** The transcription cycle consists of the three main steps of transcription initiation, transcription elongation and transcription termination. Before transcription is initiated, RNAPII is recruited to the DNA template to build the pre-initiation complex (PIC) and later an open complex is generated. Before RNAPII enters productive transcription elongation approximately 30nt downstream of the transcription start site promoter clearance is completed. Figure from (Van Lijsebettens and Grasser, 2014)

A key regulatory element which coordinates the transition through the transcription cycle is located at the C-terminus of the largest RNAPII subunit (RBP1). This carboxy-terminal domain (CTD) of RNAPII is composed of highly conserved heptapeptides with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. These residues can be subjected to multiple post-translational modifications (PTMs), like phosphorylation, glycosylation, cis-trans isomerization, and ubiquitination (Heidemann et al., 2013; Sims et al., 2011). Phosphorylation has been widely studied and is well characterized because of its important role in determining the progression of transcription and co-transcriptional processes (Lenasi and Barboric, 2013). In particular, the phosphorylation of two specific serine residues (Ser2P and Ser5P) is directly related to the coordination of transcription initiation, elongation and other RNA processing activities (Egloff and Murphy, 2008; Harlen and Churchman, 2017). Whereas in *Arabidopsis* and yeast RNAPII as part of the PIC is not modified, Ser5P modifications are enriched at the TSS what is linked with the transition into productive transcription elongation in these organisms (Hajheidari et al., 2013; Liu et al., 2004, 2009). RNAPII Ser2P modifications on the other hand spread over the gene and are enriched at the transcription end site (TES) of the gene, suggesting that this phosphorylation occurs primarily during transcription elongation and termination (Heidemann et al., 2013; Mayer et al., 2010a). The level of Ser2P increases gradually as RNAP II moves away from the promoter (Heidemann et al., 2013; Komarnitsky et al., 2000). In mammals RNAPII phosphorylated at Ser5 additionally accumulates at 5' splice sites (5'SS) of exon-intron boundaries and its density reduces as the RNAPII elongation proceeds downstream to the 3' end of the intron (Nojima et al., 2015, 2018). It has also been shown that the Ser5P phosphorylation status switches to Ser2P around 600-700 bp downstream of the TSS (Vinayachandran et al., 2018). Analysing plant native elongating transcript sequencing (pNET-seq) and global run-on sequencing (GRO-seq) data from *Arabidopsis*, an interaction between a Ser5 phosphorylated CTD and the spliceosome could be demonstrated (Zhu et al., 2018). Moreover in this study it could be demonstrated that unphosphorylated CTDs accumulate downstream the 3'SS while Ser5 phosphorylated CTDs show a peak at the 5'SS where they associate with the spliceosome. Ser2 phosphorylation on the other hand shows a peak 250 bp downstream the polyadenylation site (PAS) (Zhu et al., 2018).

## 1.2 Transcription elongation

In recent years it has been demonstrated, that transcription elongation is a dynamic and highly regulated process (Mayer et al., 2017). During this stage, RNAPII together with transcription elongation factors (TEFs) form a transcript elongation complex (TEC) that determines how the nascent RNA grows (Saldi et al., 2016). The CTD phosphorylation state plays an important role in transcription elongation. Other factors like positive transcription elongation factor P-TEFb or negative elongation factor NELF and nucleosomal barriers are additional effectors of transcription elongation (Hartzog and Fu, 2013; Peterlin and Price, 2006; Wada et al., 1998). RNAPII varies in speed during the elongation phase usually starting slowly and then accelerating within genes (Jonkers et al., 2014). Transcript elongation does not progress constantly but it consists of acceleration, backtracking, pausing and release (Nechaev et al., 2010; Weber et al., 2014). Many of the identified TEFs serve to overcome these blocks (Sims et al., 2004). Each passage that RNAPII makes along a gene is influenced by specific factors. At the beginning of transcription an exchange between GTFs and TEFs can be observed (Ehara et al., 2017).



**Figure 2. Transcription elongation factors.** A variety of transcript elongation factors as well as the CTD of RNAPII is necessary for productive transcript elongation. Some of them can directly interact with the RNAPII and help bypassing pausing steps (yellow). Others can modify histones in transcribed regions (green). Additionally ATP-dependent chromatin remodelling factors (orange) and histone chaperones (light blue) are involved in transcription elongation. Figure from (Van Lijsebettens and Grasser, 2014)

TEFs (Figure 2) are a heterogeneous group of proteins that control the efficiency of transcript elongation through the repressive chromatin (Jonkers and Lis, 2015a; Van Lijsebettens and Grasser, 2014). They can be described according to their function as modulators of RNAP II activity, modifiers of histones in transcribed regions, ATP-dependent chromatin remodelling complexes and histone chaperones (Selth et al., 2010; Sims et al., 2004). Compared to the situation in other eukaryotic systems little is known about the plant TEC, but recently it could be demonstrated that the core plant RNAPII elongation complex appears to resemble that of yeast (Antosz et al., 2017; Grasser and Grasser, 2018). The proteomic approach used by Antosz et al., showed that the *Arabidopsis* Transcription factor S-II (TFIIS), a part of the suppressor of ty 4/5 (SPT4/SPT5) complex, suppressor of ty 16 (SPT6), RNAPII-associated factor 1 (PAF1-C), and the histone chaperone complex facilitates chromatin transcription (FACT) associate with elongating RNAPII similar to yeast (Krogan et al., 2002). Genetic and biochemical studies further revealed that various TEFs regulate plant growth and development (Antosz et al., 2017; Grasser and Grasser, 2018; Van Lijsebettens and Grasser, 2014). Since the TEFs TFIIS, IWS and SPT4 were analyzed in this study, in the following these factors will be described in more detail.

## TFIIS

Transcription factor S-II (TFIIS) belongs to the group of TEFs that influence the transcriptional properties of elongating RNAPII (Van Lijsebettens and Grasser, 2014). It positively regulates productive transcript elongation by stimulating hydrolytic RNA cleavage activity of RNAPII reducing the time of pausing and stimulating the rate of RNAP II transcriptional elongation (Antosz et al., 2020; Wu et al., 2015). TFIIS plays a crucial role in regulating transcript elongation (Fish and Kane, 2002) and its activity has been demonstrated to be necessary for the efficient release of paused RNAPII from promoter proximal sites (Adelman et al., 2005; Saunders et al., 2006). TFIIS further induces

extensive structural changes in RNAPII and facilitates realignment of the RNA in the active site for catalyses (Kettenberger et al., 2003). In *Arabidopsis* the loss of TFIIIS does not result in a severe plant growth phenotype (Grasser et al., 2009; Mortensen and Grasser, 2014) but *tfl1s* plants are affected at the end of the maturation phase with seeds showing reduced dormancy. At later developmental stages this results in the misregulation of a variety of genes (Liu et al., 2011). The TFIIIS gene is expressed seed specifically and is also used as a quantitative locus for seed dormancy in *Arabidopsis* (Bentsink et al., 2006).

## IWS1

IWS1, (Snp1 in yeast), is a TEF that most likely modulates the transition to productive transcript elongation (Pujari et al., 2010). It interacts transcription elongation factor SPT6 (Li et al., 2010) and chromatin remodelling complex SWI/SNF to induce *CYC1* gene expression (Krogan et al., 2002; Zhang et al., 2008) supporting a role in transcription elongation. SPT6 on the other hand can form a complex with the Ser2P modified CTD of RNAPII (Vos et al., 2018). SPT6/IWS1 interact with the methyltransferase SETD2 (Oqani et al., 2019; Yoh et al., 2007). In *Arabidopsis* IWS1 is part of a group of TEFs that modulate histone modifications (Subtil-Rodríguez and Reyes, 2011; Widiez et al., 2011). It is also involved in modulating BR-regulated gene expression and it directly interacts with the histone lysine methyltransferase SDG8 (Li et al., 2010; Wang et al., 2014). Recently it was supposed that IWS1 directly regulates pre-mRNA splicing by interacting with the spliceosome (Kanno et al., 2020).

## SPT4

SPT4 is part of the suppressor of ty 4/5 (SPT4/SPT5) heterodimeric complex which couples chromatin modification states and RNA processing to transcription elongation (Hartzog and Fu, 2013). Through the modulation of pausing or arrest during transcription, this complex can regulate the elongation of RNAPII (Van Lijsebettens and Grasser, 2014; Vos et al., 2018). An accumulation of unspliced pre-mRNAs and several intron-containing RNAs in *spt* mutants further indicates a possible role of SPT4/5 in splicing (Lindstrom et al., 2003). In *Arabidopsis*, downregulation of *SPT4/SPT5* expression affects vegetative and reproductive development and results in severe growth defects (Dürr et al., 2014a). Plants depleted of SPT4 also show a misregulation of certain auxin-related genes (Dürr et al., 2014). SPT4/SPT5 further co-immunopurify and genetically interact with RNAPII and the Paf1 complex (Antosz et al., 2017a). Additionally, in yeast, it has been shown that SPT4 is involved in the regulation of H3K36me3 mark deposition (Lee et al., 2017).

## 1.3 Co-transcriptional splicing of mRNAs

The generation of the mRNA takes place in a nuclear environment where several processes important for mRNA maturation occur in a remarkable spatial and temporal proximity. These processes occur co-transcriptionally and comprise the mRNA processing events 5' capping, splicing and polyadenylation (Bentley, 2014; Fong et al., 2014; Saldi et al., 2016).

The splicing process itself is carried out by the spliceosome. This huge machinery (~ 3 mDa in human cells) consists of 5 snRNAs and nearly 300 proteins (Galej, 2018; Jurica and Moore, 2003; Will and

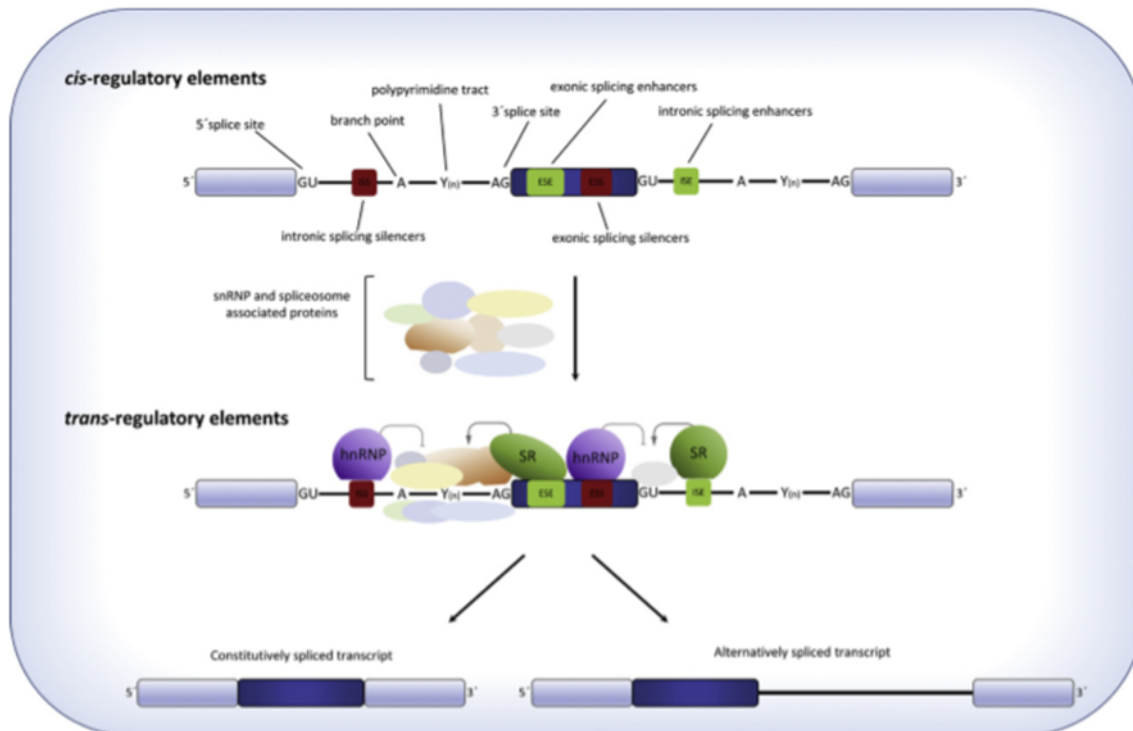


Lührmann, 2011). It precisely removes introns and joins exons together in order to generate a single mRNA molecule (Jangi and Sharp, 2014; Reddy et al., 2013). In higher eukaryotes there are two types of spliceosomes. The major U2-type spliceosome, containing U1, U2, U4, U5 and U6 snRNPs, catalyses the removal of the majority of introns with canonical splice sites. The minor U-12 type spliceosome, containing U11, U12, U4atac, U5 and U6atac snRNPs, on the other hand recognises only a small number of introns and uncanonical splice sites (Reddy, 2007b). In plants most of the splicosomal proteins are conserved, suggesting that the composition is similar to that of animals but the plant spliceosome is still not well characterized (Albaqami and Reddy, 2018).

Splicing is regulated by trans-acting proteins (repressors and activators) and corresponding cis-acting regulatory sites (silencers and enhancers) on the pre-mRNA. Two major classes of universal splicing regulatory factors are constituted by the serine/arginine-rich (SR) proteins and members of the heterogeneous nuclear ribonucleoprotein (hnRNP) protein family (Figure 3) (Naftelberg et al., 2015). SR proteins are composed of one or two RNA recognition motifs (RRMs) at the N-terminus and a C-terminal domain which comprises several arginine/serine-rich (RS) motifs (Reddy and Shad Ali, 2011). SR proteins are essential for both constitutive and alternative splicing (Bourgeois et al., 2004). They can recognise specific sequence motifs located in exons or introns and influence splice site selection (Jeong, 2017). The RRM motifs recognise *cis*-acting sequences whereas the RS domain is important for protein-protein interactions to recruit other proteins of the splicing machinery and to contact the pre mRNA branch point (Jeong, 2017; Morton et al., 2019; Reddy, 2007a). It has also been shown that SR proteins form a complex network of direct interactions between each other and with other spliceosomal proteins (Graveley, 2001; Lorković et al., 2005). SR proteins are mainly splicing enhancers to which splicing activator proteins bind, increasing the probability that a nearby site will be used as a splice junction (Reddy and Shad Ali, 2011). They can function as both, intronic splicing enhancers (ISE) or exonic splicing enhancers (ESE) (Morton et al., 2019). SR proteins can also act in auto- and cross- regulatory negative feedback loops by triggering splicing of their own pre-mRNA or that of related proteins (Barbazuk et al., 2008; Simpson et al., 2008; Stauffer et al., 2010). In *Arabidopsis* the number of SR proteins is higher than in mammalian systems (Morton et al., 2019).

Proteins of the hnRNP family mainly repress splicing by either antagonizing the recognition of splice sites or by interfering with the binding of proteins bound to enhancers (Martinez-Contreras et al., 2007; Morton et al., 2019). They can function as both intronic splicing silencers (ISS) or exonic splicing silencers (ESS) (Naftelberg et al., 2015). Beside a role as splicing repressors, some hnRNP proteins can also function as splicing activators to maintain the transcriptome integrity (Das et al., 2019).

It has become evident that the mode of action of splicing regulatory factors is often related to the binding position within the pre-mRNA and by the interaction of multiple components of the splicing machinery and additional regulatory factors (Erkelenz et al., 2013; Ruhl et al., 2012).



**Figure 3. Splicing is regulated by *cis*- and *trans*-regulatory elements.** Several SR and RNA binding proteins take part in the early stages of spliceosome assembly. *Trans*-factors like SR-proteins interact directly with *cis*-exonic splicing regulators (ESR). The SR proteins recognise specific sequences at 5' and 3' splice sites and bridge spliceosomal components. hnRNP proteins bind usually intronic splice regions and are splicing repressors. Figure from (Morton et al., 2019)

## 1.4 mRNA splicing in plants

Like in animals, also in plants the majority of genes (~ 80 %) contain non-coding sequences that are processed to generate mature mRNAs, (Labadorf et al., 2010; Reddy, 2007a). Bioinformatic analyses using sequence similarity further support the idea that the core plant spliceosome is conserved (Albaqami and Reddy, 2018; Reddy et al., 2013). Furthermore sequences in 5' and 3' splice sites, polypyrimidine tract and branch point sequence (BPS) are similar in plants and animals indicating a similar basic mechanisms of intron processing in plants (Reddy, 2007a). Other findings indicate that also a plant-specific splicing regulatory mechanism exists. For instance it was shown that animal introns cannot be properly processed and spliced in plant systems (Albaqami and Reddy, 2018; Haseloff et al., 1997). Although *Arabidopsis* and human cells have similar 5' and 3' splice sites, the average length and nucleotide composition of genes and introns is different (Reddy, 2007a). Introns also differ in the composition of U and UA bases what was shown to be fundamental for the recognition of the splice site and for efficient splicing (Brown and Simpson, 1998; Simpson et al., 2004). This indicates that some events involved in the splice site recognition are unique in plants. Additionally, a higher presence of splicing regulators in plants like SR proteins, SR protein kinases and hnRNP proteins suggests differences in sequence requirement between plant and animal splice site recognition (Albaqami and Reddy, 2018; Brown and Simpson, 1998; Morton et al., 2019).

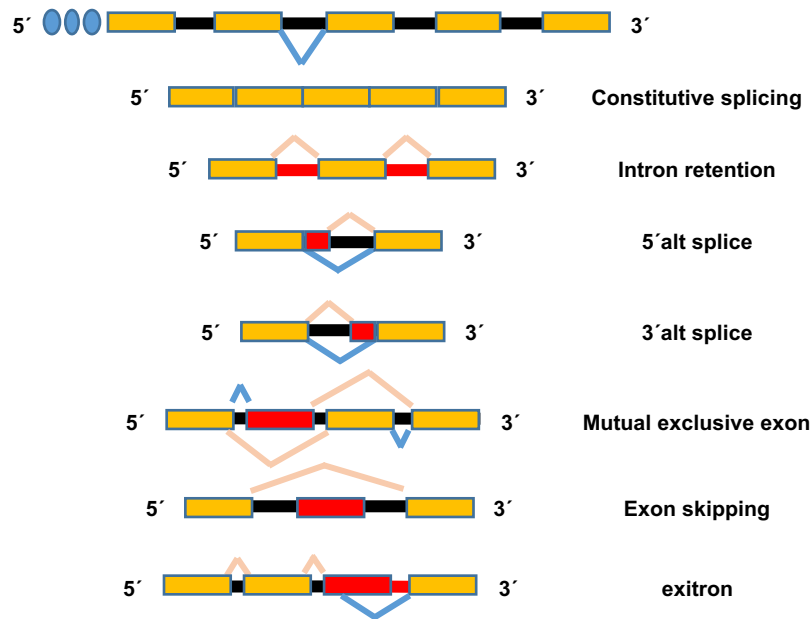
However, many aspects of pre-mRNA splicing in plants are still undefined and need further elucidation (Reddy et al., 2013).

## 1.5 Alternative splicing

Splicing can also increase the transcriptome and proteome complexity within cells and tissues. After the discovery of exons and introns in adenovirus 2 genes (Berget et al., 1977; Chow et al., 1977), it was suggested that more combinations of exons could be joined together to produce different isoforms of mRNAs, a process called alternative splicing (AS). AS may lead to phenotypic variation from an identical genotype what might explain the variable appearance of organisms with identical genetic information (Reddy, 2007a). The natural organization of eukaryotic genes in coding regions interrupted by non-coding introns seems to have evolutionary advantages. The presence of a splicing-related mechanism behind the higher complexity of vertebrates compared to invertebrates is supported by the fact that there is not a huge difference in the amount of protein-coding genes in vertebrates and invertebrates, but the number of spliced isoforms per gene is higher in vertebrates (Naftelberg et al., 2015). The importance of AS in multicellular eukaryotes could clearly be demonstrated over the last years and it was shown that it has the capacity to create enormous proteome and transcriptome complexity (Chaudhary et al., 2019a; Reddy, 2007a).

There are several common types of alternative splicing events (Figure 4):

- i) intron retention: an intron can be included in the transcript, generating an mRNA of a different size (Naftelberg et al., 2015)
- ii) alternative 5' or 3' splice site: mRNAs of different sizes are produced depending on the usage of a proximal or distal 5' or 3' splice site (Naftelberg et al., 2015)
- iii) mutually exclusively exon: adjacent exons are spliced in such a way that only one is included at the time of mRNA formation (Naftelberg et al., 2015)
- iv) cassette exons (exon skipping): an exon can be included or excluded from the transcript; (Naftelberg et al., 2015)
- v) exitron (EI): alternatively spliced internal regions with both features of introns and exons (Marquez et al., 2015).



**Figure 4. Schematic illustration of AS events.** Yellow boxes represent constitutive exons; red boxes alternative regions; black lines introns. Pink lines show alternative splicing events and light blue lines constitutive splicing events.

## 1.6 Alternative splicing: humans vs. plants

In human cells ~ 95% of all intron-containing genes are estimated to be alternatively spliced (Pan et al., 2008) whereas in *Arabidopsis* up to 70% of genes undergo alternative splicing (Marquez et al., 2012; Syed et al., 2012; Zhang et al., 2017). Even though AS is more prevalent and better described in human cells, studies provide evidence that AS is also an important layer of gene expression regulation in plants. In both systems AS regulates essential functions such as autophagy, apoptosis, protein localization, enzymatic activities and interaction with ligands, transcription factors activity and mRNA abundance in human cells and light regulation, abiotic and biotic stress response and circadian clock in plants (Chaudhary et al., 2019a; Godoy Herz et al., 2019a).

The contribution of AS toward protein diversity is well described in human cells but there are also studies that suggest that AS may not be a key contributor to protein diversity (Gallego-Paez et al., 2017; Liu et al., 2017). In *Arabidopsis* the role of AS in the expansion of protein diversity is less documented but it was suggested that unique events of AS may contribute to protein diversity (Marquez et al., 2012).

In human cells exon skipping events, that are likely to be translated, are the most prevalent AS events, while in *Arabidopsis* intron retention is the predominant mode of AS (Chaudhary et al., 2019a). Resulting transcripts with retained introns are often degraded by the nonsense-mediated decay (NMD) pathway suggesting that in plants AS often allows fine-tuning of gene expression by changing the ratios of productive and unproductive variants (Chaudhary et al., 2019b).

## 1.7 Coupling between alternative splicing and transcription

The first prominent evidence for the coupling of transcription and splicing was shown more than thirty years ago. Electron microscopy on *Drosophila* embryos revealed the presence of splicing loops in

nascent transcripts (Beyer and Osheim, 1988). Over the last decade genome-wide deep sequencing approaches confirmed that in eukaryotic cells splicing occurs mainly during transcription (Drexler et al., 2019; Nojima et al., 2015, 2018; Tilgner et al., 2012). The kinetics of co-transcriptional splicing is well described in yeast (Wallace and Beggs, 2017). Splicing has been shown to occur within a short kinetic window following RNAPII elongation and pausing could be observed over the 3'SS of downstream exons in budding yeast (Carrillo Oesterreich et al., 2016). A mathematic model further suggests that pausing is fundamental to allow co-transcriptional splicing and that this mechanism might be evolutionary conserved in *S. cerevisiae* (Carrillo Oesterreich et al., 2010). In mammals in contrast, co-transcriptional splicing occurs after RNAPII transcribes several kilobases of pre-mRNA whereas the interplay of regulatory proteins that bind to specific RNA sequences (splicing enhancers and splicing silencers) dictate splicing (Drexler et al., 2019; Nojima et al., 2018). It was also shown in human and plants that for the majority of genes chromatin environment, nucleosome occupancy and RNAPII processivity have a strong influence on the splicing outcome (Jabre et al., 2019; Zhu et al., 2018; Saldi et al., 2016). The key regulatory elements for coupling transcription and splicing are the CTD of RNAPII and the chromatin environment, in which both transcription and splicing take place (Naftelberg et al., 2015).

Several studies also revealed a coupling of transcription and AS (De La Mata et al., 2003; Nogués et al., 2002). Interestingly, compared to the co-transcriptional catalysis of constitutive splicing, the alternative splicing process itself can also occur post-transcriptionally, at least in a group of studied events (Vargas et al., 2011). Besides promoters, also transcription factors, co-activators, transcriptional enhancers, chromatin remodelling factors and factors affecting chromatin structure were shown to influence the alternative splicing decision (Naftelberg et al., 2015).

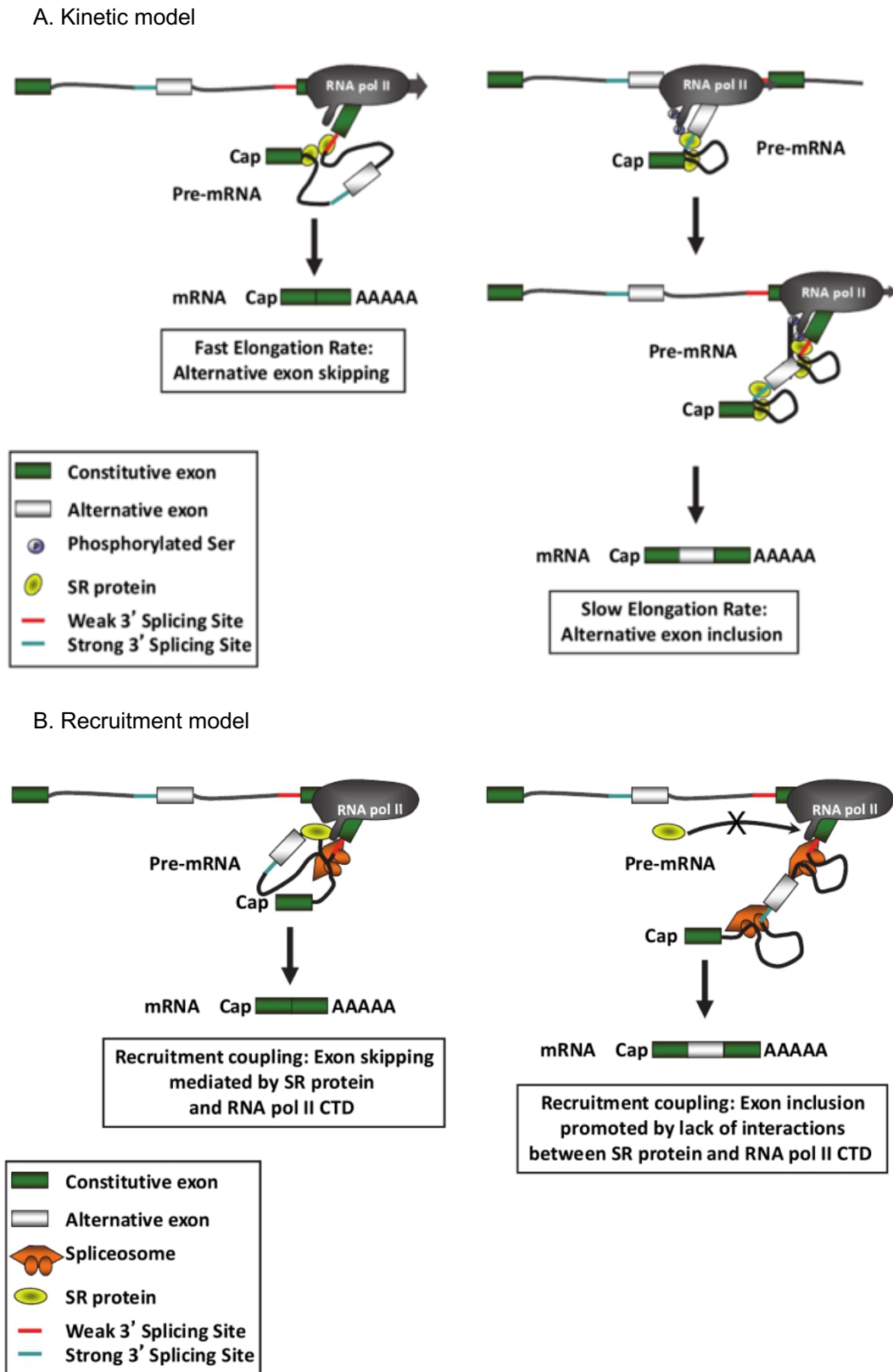
## 1.8 Kinetic coupling vs. recruitment coupling

Two mechanisms have been proposed to explain the coupling between AS and transcription: kinetic coupling and recruitment coupling (Fong et al., 2014; Naftelberg et al., 2015). The kinetic coupling model is based on how the RNAPII elongation rate affects the choice of alternative splice sites (Fong et al., 2014). Evidence that the transcript elongation rate regulates splice site selection was given by analysing a RNAPII mutant which encodes a point mutation that produces a slow transcription rate in human cells (De La Mata et al., 2003). The theory of kinetic coupling between AS and transcription is based on the observation that a slow RNAPII transcription rate results in the recognition also of weak splice sites and the inclusion of alternative exons. If RNAPII elongation rate is fast on the other hand, the strong splice sites are recognized by the splicing machinery more efficiently, resulting in the skipping of alternative exons (Figure 5). However, in some cases a slow RNAPII elongation rate also promotes exon skipping in a mechanism that involves the recruitment of negative factors to splice sites (Fong et al., 2014; Godoy Herz et al., 2019a).

The recruitment model focuses on the recruitment of splicing factors to the TEC by an interaction with the CTD of RNAPII and/or the chromatin status. A dynamic phosphorylation of the CTD regulates the passage through the transcription cycle. Whereas phosphorylation of the CTD repeats on Ser5 residues is crucial for the recruitment of capping enzymes at 5' ends of genes, cleavage/polyadenylation factors are recruited to the 3' ends by Ser2 phosphorylated residues of the CTD (Saldi et al., 2016). Additionally several factors involved in splicing and alternative splicing were shown to be recruited to the transcription machinery by CTD but it is currently not clear if these interactions are mediated by RNAPII or nascent mRNA (Das et al., 2007; Gu et al., 2013; De La Mata and Kornblihtt, 2006; Naftelberg et al., 2015; Saldi et al., 2016).

Although kinetic coupling and recruitment coupling describe autonomous models, these are not mutually exclusive. The RNAP II elongation rate is directly linked to the recruitment of specific processing factors. At the same time factors that control the elongation rate are differentially recruited to the TEC (Brody et al., 2011a; Jimeno-González et al., 2015; Saldi et al., 2016). In human cells it could be shown that RNAPII pauses at 3' and 5' splice sites that flank exons at what RNAPII CTD is hyperphosphorylated at Ser5 residues. The release from splice site associated pauses goes along with an increase in Ser2 CTD phosphorylation what might be stimulated by splicing factor U2AF65 that associates with the TEC (Mayer et al., 2015; Nojima et al., 2015; Saldi et al., 2016; Újvári and Luse, 2004).

The fact that recruitment of factors to RNAPII and kinetic coupling is likely interdependent was further indicated by analysis of the transcription factor TFIIS. TFIIS can rescue backtracked RNAPII and it reduces the duration of RNAPII pausing and accelerates elongation (Saldi et al., 2016). Depletion of TFIIS accelerates pausing what results in enhanced exon inclusion suggesting that pausing of backtracked RNAPII provides the possibility to complete some splicing reactions (Dolata et al., 2015a; Howe et al., 2003; Saldi et al., 2016).



**Figure 5. AS can be explained by the kinetic model and the recruitment model.** A) Kinetic model: the RNAPII elongation rate influences the inclusion or skipping of exons. A fast elongation rate results in skipping of weak splice sites whereas a slow elongation results in the recognition also of weak splice sites leading to exon inclusion. B) Recruitment model: a dynamic phosphorylation status of the CTD leads to the recruitment of specific splicing factors (SR proteins) resulting in exon skipping or exon inclusion. Figure from (Hall and Georgel, 2011)

## 1.9 The chromatin structure as a regulator of AS

Apart from the RNAPII transcription complex also the chromatin structure plays a key role as an effector of splicing. Nucleosomes and histone modification are not randomly distributed between exon and intron (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009). Nucleosomes for instance act as a barrier for transcription and are mostly more tightly packed on exons than on introns (Huang et al., 2012; Schwartz et al., 2009). Since consequently a higher RNAPII occupancy and a slower elongation rate were determined in exons, kinetic coupling suggests that pausing at nucleosomes may influence alternative splicing decisions (Saldi et al., 2016). Accordingly, in histone-depleted cells, chromatin accessibility and RNAPII elongation rate was increased and a variety of transcripts showed altered alternative splicing (Jimeno-González et al., 2015).

The nucleosome barrier for transcript elongation is *in vivo* overcome by the help of histone chaperones and chromatin remodelers. The key chaperones of histone dynamics FACT (facilitates chromatin transcription) and SPT6 both increase the rate of RNAPII elongation through chromatin. Also the chromatin remodelers SWI/SNF and CHD1 promote RNAPII progress through nucleosomes (Saldi et al., 2016). Additionally, these factors interact with splicing factors and their depletion results in altered AS. Thus kinetic coupling implies histone chaperones and chromatin remodelers but these factors might also be involved in the regulation of AS by recruiting factors involved in mRNA processing (Saldi et al., 2016).

Exons and introns do not only differ in nucleosome occupancy, they also display a different density of covalent histone marks (Gelfman et al., 2013; Saldi et al., 2016). These post-translational modifications (PTMs) of histones like H3 lysine acetylation, methylation, serine phosphorylation, ADP ribosylation and ubiquitination are correlated with nucleosome occupancy. Thus they may influence RNAP II elongation speed and splicing factor recruitment resulting in different splicing outcomes (Luco et al., 2010; Roudier et al., 2011; Zhou et al., 2014). The most common PTMs are acetylation and methylation which have been classified historically as activating or repressing (Saldi et al., 2016). More recent studies using genome-wide chromatin immunoprecipitation-sequencing (ChIP-seq) demonstrated that the impact of a given PTM on transcription is more context dependent (Saldi et al., 2016). Several histone marks associated with actively transcribed genes including H3K36me3, H2BK120 monoubiquitylation (H2Bub1) and H3, H4 acetylation were shown to be implicated also in the control of splicing (Zhou et al., 2014).

## 1.10 The role of H3K36me3, MRG proteins and PTB in the regulation of alternative splicing

Several studies analysing ChIP-seq data identified a pronounced enrichment of the H3K36me3 mark on exons, indicating that exon/intron boundaries are marked by transitions between regions enriched in and depleted of H3K36me3 (Andersson et al., 2009; Saldi et al., 2016; Schwartz et al., 2009). Also other PTMs like H3K4me3, H3K27me2, H3K27me3, H3K79me1 and H3K9me3 have been reported to be differentially enriched or depleted in exons but to a lesser extent (Andersson et al., 2009; Saldi et al., 2016; Schwartz et al., 2009). The role of H3K36me3 in splicing has been shown in several studies. Levels of H3K36me3 are lower at the promoter and higher in the transcribed regions of genes (Wagner and Carpenter, 2012). H3K36me3 is usually associated with actively transcribed regions and with a concomitant increase in nucleosome occupancy (Kolasinska-Zwierz



et al., 2009; Spies et al., 2009). Interestingly, H3K36me3 marks are associated more significantly with constitutive exons than with alternative exons suggesting a relationship with *cis*-splicing (Allemand et al., 2008; Kolasinska-Zwierz et al., 2009). Although H3K36me3 could modulate co-transcriptional splicing through kinetic coupling, well-documented examples demonstrated that H3K36me3 dependent regulation of splicing involve chromatin-splicing adaptor systems (Luco et al., 2010; Sorenson et al., 2016).

In mammalian cells it has been shown that the trimethylation of H3K36 is mediated by the methyltransferase SETD2 in a mechanism that also involves the TEF IWS1 (Sanidas et al., 2014; Lee and Blenis, 2014).

In *Arabidopsis* 43 proteins contain a SET domain which is characteristic for the human methyltransferase SETD2. The *Arabidopsis* H3K36-methyltransferase SET DOMAIN GROUP 8 (SDG8) directly interacts with IWS1 and the depletion of SDG8 results in a reduction of H3K36me2 and H3K36me3 indicating that SDG8 methylates H3K36 modifications in *Arabidopsis* (Wang et al., 2014; Zhao et al., 2005).

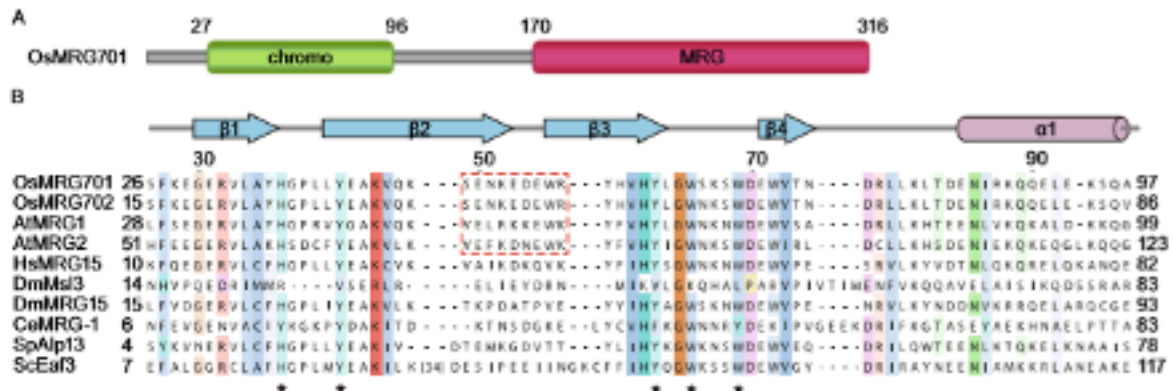
The histone mark subsequently recruits the MORF-related gene on chromosome 15 (MRG15) chromatin binding protein (Sanidas et al., 2014). An MRG15-like protein was first discovered in yeast (Eaf3) as a reader of the H3K36me3 mark (Carrozza et al., 2005; Joshi and Struhl, 2005). MRG15-like proteins have been shown to be conserved containing an N-terminal chromodomain that binds to an H3K36me3 mark and a C-terminal MRG domain that displays sequence similarity with the human Mortality factor on chromosome 4 (MORF4) which is involved in mediating protein-protein interaction (Figure 6) (Bertram and Pereira-Smith, 2001). MRG15 chromatin reader proteins further recruit the polypyrimidine tract-binding protein (PTB) (Sanidas et al., 2014). The binding of the splicing factor PTB to an intronic silencer element can influence the splicing of different isoforms as demonstrated for the human fibroblast growth factor receptor 2 gene (FGFR2) (Luco et al., 2010). The mechanism how PTB controls AS includes competing with U2 auxiliary factor 65 in binding to the pre-mRNA (Sauliere et al., 2006), looping of RNA regions and interfering with splicing factor interactions required for exon or intron definition (Izquierdo et al., 2005; Sharma et al., 2005).

Moreover, two MRG15-like candidates can be identified in *Arabidopsis* based on sequence similarity. MRG1 and MRG2 share ~49% identity at the amino acid sequence level. They have been shown to promote photoperiodic regulation of flowering time in *Arabidopsis* by regulating the expression of two specific genes involved in flowering time control with opposite functions: the FLOWERING LOCUS T (FT) and the FLOWERING LOCUS C (FLC) (Bu et al., 2014; Xu et al., 2014). The regulatory function of MRG1/2 is dependent on the binding to H3K36me3 which is mediated by the respective chromodomain. Additionally, an interaction with H4-specific acetyltransferases (HAM1 and HAM2) was discovered resulting in an increase of H4 acetylation levels demonstrating that MRG proteins act as a bridge between H3K36 methylation and H4 acetylation in *Arabidopsis* (Xu et al., 2014).

Furthermore it could be shown that MRG1/2 proteins bind PHYTOCHROME-INTERACTING FACTOR7 (PIF7) and are thereby involved in the epigenetic regulation of hypocotyl elongation and gene transcription (Peng et al., 2018).

Besides orthologues of the methyltransferase and the respective chromatin reader protein also three candidates with sequence similarities to the human splicing factor PTB can be identified in *Arabidopsis*. Among those, PTB1 and PTB2 are involved in regulating AS of genes involved in the control of several fundamental developmental processes (Ruhl et al., 2012a).

A role of the H3K36me3 in transcription elongation has also been hypothesized for plants since this histone mark is also enriched in transcribed regions (Roudier et al., 2011). Although it seems that the key players of the H3K36me3-mediated regulation of AS are conserved in *Arabidopsis*, it is not known if this mechanism is also conserved.



**Figure 6. MRG15-like proteins.** A) MRG proteins display an N-terminal chromo domain and a C-terminal MRG domain. B) Sequence alignment of MRG15-like proteins in *Oryza sativa*, *Arabidopsis thaliana*, *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Schizosaccharomyces pombe* and *Schizosaccharomyces cerevisiae*. Amino acids are shaded as a function of conservation with a 30% conservation visibility cutoff. Secondary structural elements of MRG701CD are displayed above the sequence. Figure from (Liu et al., 2016)

In summary a variety of studies performed over the last decade point to two possible mechanisms by which chromatin may act as an effector of AS. A) Chromatin can affect transcript elongation through kinetic coupling which results in localized transcriptional pausing leading to a possible shift in splicing reactions. B) Chromatin may act as a scaffold to recruit splicing factors (Saldi et al., 2016).

These studies also demonstrated that the regulation of AS is quite complex because on the one hand alternative splicing is regulated by both *cis*- and *trans*-regulatory elements and on the other hand it is coupled to transcription and factors that regulate transcription are also affected by alternative splicing (Fong et al., 2014; Wu et al., 2015; Zhu et al., 2020; Saldi et al., 2016).

# Aim of the project

## The effect of transcription elongation factors on alternative splicing

In eukaryotes alternative splicing increases protein diversity. Splicing occurs mainly co-transcriptionally and the RNAPII processivity was shown to play a key role in regulating the splicing outcome. In this study the influence of transcription elongation factors on AS will be analyzed. Therefor RNAseq will be performed in different transcription elongation factor mutants. Bioinformatic analyses of RNAseq data should reveal genes that are differentially alternatively spliced in the different transcription elongation mutants. Additionally, the characteristics of AS events will be determined in more detail. To confirm the RNAseq data, RT-PCR on selected differentially alternatively spliced will be conducted. To investigate the impact of TEFs on RNAPII elongation globally, ChIPseq experiments immunoprecipitating Ser2P and Ser5P RNAPII will be performed.

## The chromatin status and alternative splicing

Besides the RNAPII elongation rate also chromatin adapter complexes can influence splicing decisions. In human cells H3K36me3 histone marks can recruit the MRG15 chromatin reader protein which in turn recruits the splicing regulator PTB. In *Arabidopsis* two orthologues of MRG15 are described and a possible splicing regulatory pathway through an H3K36me3 – MRGs chromatin adapter complexes will be analysed in this study. To identify putative MRG interactors, a proteomic approach after affinity purification using MRG1 and MRG2 bait proteins will be applied and it will be screened for co-purifying splicing factors and methyltransferases. A possible protein-protein interaction between co-purifying will be additionally tested by yeast-two hybrid and FRET experiments. In human cells the methyltransferase SETD2 (SDG8 in *Arabidopsis*) which deposits the H3K36me3 histone marks is recruited to the chromatin by the transcription elongation factor IWS1. Thus it will be examined if also in *Arabidopsis* IWS1 is involved in the regulation of a putative H3K36me3 – MRGs chromatin adapter complex. A mutant plant line will be generated lacking IWS1, MRG1 and MRG2 and by phenotyping it will be tested if IWS1 and MRGs may function in the same regulatory pathway. Finally the impact of IWS1 on global H3K36me3 levels will be investigated by western blotting and ChIPseq experiments.



## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

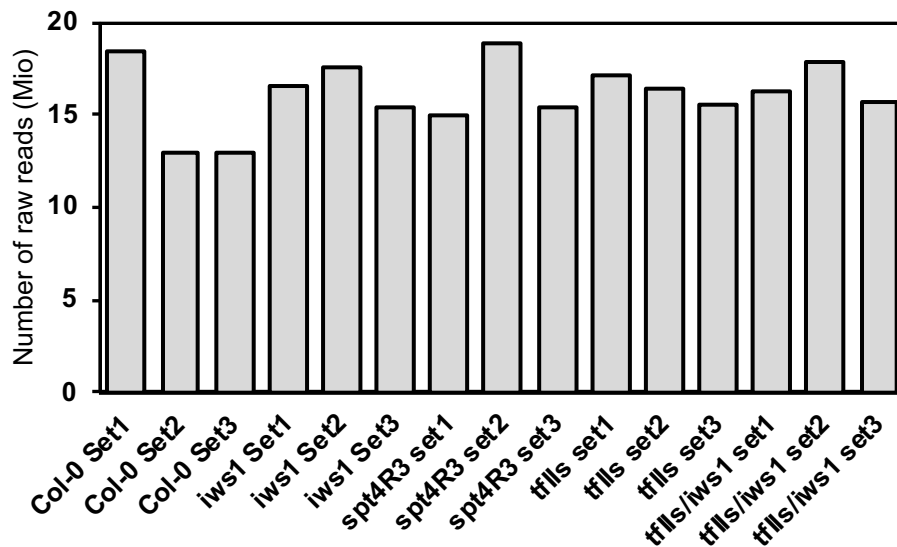
### 2.1 The *tef* mutants *spt4R3*, *tflls*, *iws1* and *tflls/iws1* show alternative splicing defects

The study analysing a putative interplay between transcription and alternative splicing in *Arabidopsis thaliana* was started in our work group by the three bachelor students Laura de Jonge, Isabel von der Decken and Raphael Rager under the supervision of Dr. Marion Grasser. Laura de Jonge screened different *tef* mutant T-DNA lines and analysed a possible coupling between transcription and splicing (BSc Laura de Jonge, 2013). Isabel von der Decken (BSc Isabel von der Decken, 2014) tested different mRNA splicing isoforms of three SR genes and she detected altered splicing via sequencing. Raphael Rager (BSc Raphael Rager, 2014) analysed the “seven indicator genes” (Palusa and Reddy, 2010; Palusa et al., 2007) what indicated that the TEFs TFIIIS, IWS1 and SPT4 are involved in the regulation of alternative splicing. He also performed the first bioanalyzer analysis. These analyses revealed reproducibly that AS events are altered in the *tef* mutant lines *tflls*, *iws1*, *spt4R3* as well as in the double mutant line *tflls/iws1*. To analyse splicing pattern globally, an RNA-seq analysis was performed using mRNAs from these four mutant plant lines and transcriptomes were compared regarding variation in splice variants compared to Col-0.

### 2.2 RNAseq data analysis reveals differentially alternatively spliced genes in *tef* mutants

Transcriptomes of *Arabidopsis tef* mutant seedlings grown for 14 days under long day conditions were analysed to investigate the potential impact of transcription elongation on alternative splicing. RNA-sequencing allows the assessment of differential expression of genes through quantification of transcripts across a broad dynamic range. The influence of mutations in the elongation factors TFIIIS, IWS1 and SPT4 on alternative splicing were analysed in the homozygous T-DNA insertion mutant lines *tflls-1-/-* (*tflls*) and *iws1-/-* (*iws1*), as well as in the RNAi line *spt4R3* (*spt4R3*). In view of the finding of significant AS differences in *iws1* and *tflls* mutant plants a double mutant *tflls/iws1* was generated to test a possible interaction of the TEFs TFIIIS and IWS1 in modulating mRNA splicing. Using an RNA sequencing approach it was intended to identify splicing targets regulated by TEFs and to define the potential global impact of transcription on AS. To identify and quantify splicing events, the resulting raw reads were mapped to the *Arabidopsis* genome (TAIR10 version) resulting in approximately 160 mio reads per library (Figure 7). The obtained data of high quality allowed rigorous statistical analyses of detected splicing events. The results showed consistence between the replicates and reproducibility between the mutants. AS events were extracted from TAIR10 and complemented by unannotated events that were identified in the RNAseq data. 33.416 genes out of 33,602 genes annotated in TAIR10 were expressed in at least one sample. Library preparation and RNAseq were performed at the Genomics Core Facility “KFB - Center of Excellence for Fluorescent Bioanalytics” (University of Regensburg, Regensburg, Germany; [www.kfb-regensburg.de](http://www.kfb-regensburg.de)). The RNAseq data were analyzed by the bioinformatics Dr. A. Kahles, Prof. Dr. G. Rättsch (ETH, Zurich) and Prof. Dr. A. Wachter (University of Mainz).

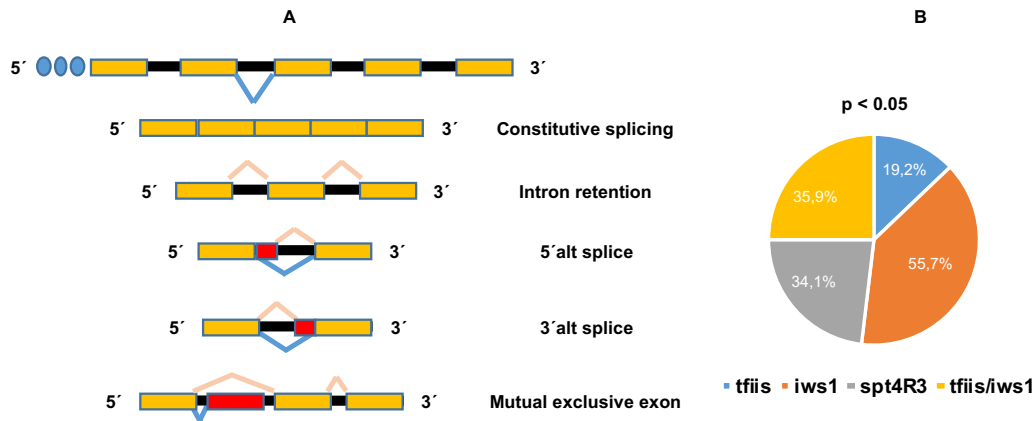
2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



**Figure 7.** Number of raw reads generated by Illumina sequencing in the analysed plant lines Col-0, *iws1*, *spt4R3*, *tflls* and *tflls/iws1*. Three biological replicates were sequenced for every genotype (Set1-3).

A schematic representation (Figure 8 A) shows the most relevant alternative splicing events identified in the data: intron retention (IR), 5' alternative splice site (5'SS), 3' alternative splice site (3'SS) and mutual exclusive exon. Of all expressed alternatively spliced genes 167 showed significantly altered AS in the mutants analyzed compared to Col-0 (Supplementary table 1). Among those 35.9% showed significant alternative splicing changes in the double mutant genotype *tflls/iws1*, 55.7% in *iws1*, 19.2% in *tflls* and 34.1% in *spt4R3* (Figure 8 B). The identification of some 'seven indicator genes' that were used in the first screening experiments and some genes encoding SR proteins as differentially alternative spliced in the *tef* mutant plant lines (Supplementary table 1) further confirmed the quality of the RNAseq data.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



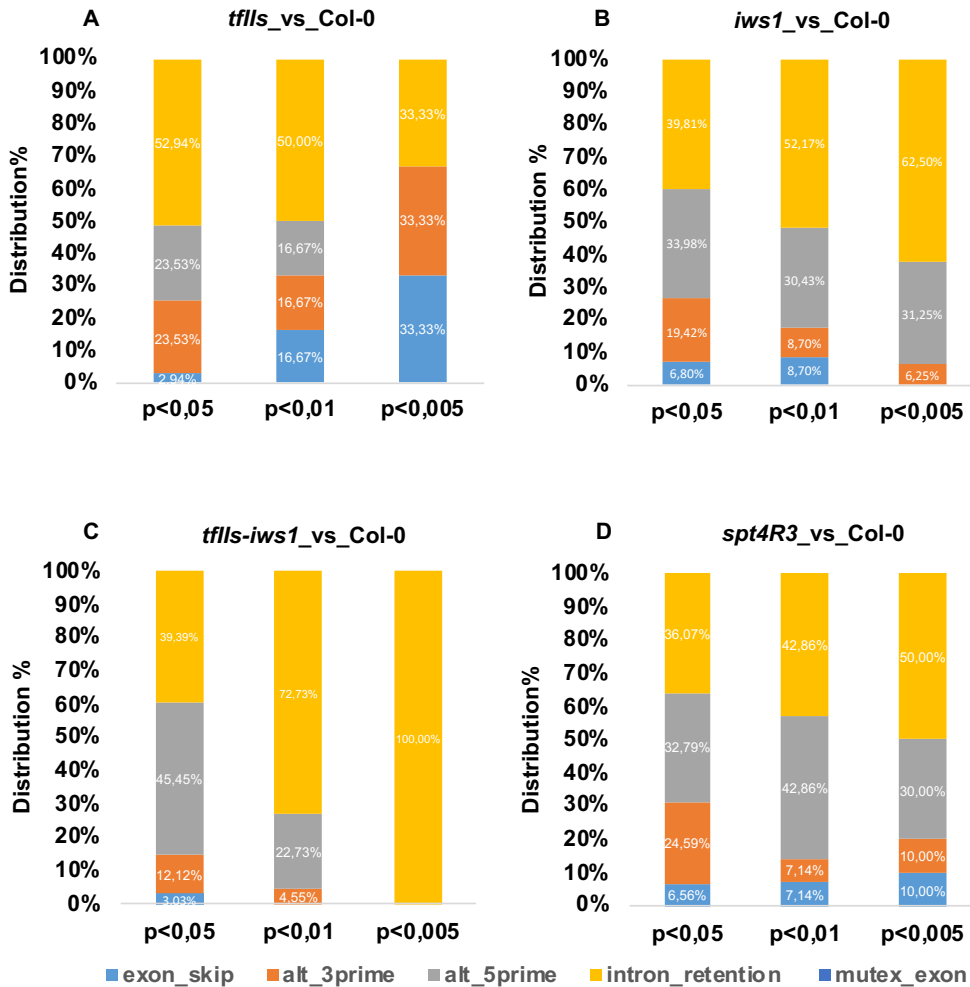
**Figure 8. AS events identified by RNAseq.** A) Schematic illustration of the most relevant alternative splicing events found in the data. Yellow boxes represent constitutive exons; red boxes alternative regions; black lines introns. (B) Percentage of genes identified in RNAseq data that showed altered alternative splicing events in the *tef* mutants compared to Col-0. Only genes with significant changes  $p < 0.05$  are represented.

When analyzing the specific splicing events that were differentially affected in the different *tef* mutant lines it could be observed that IR is the most prominent represented AS event (Figure 9). This is in line with previous studies in plants (Drechsel et al., 2013a; Filichkin et al., 2015; Marshall et al., 2016; Seo et al., 2011) whereas exon-skipping is the prevalent AS event detected in human cells (Wang et al., 2008). When applying the most stringent filter to identify genes with altered alternative splicing events ( $p < 0,005$ ), in the mutant lines *iws1* and *spt4R3* the majority of all genes differentially affected in alternative splicing belong to this class (Figure 9 B,D, 62.50% and 50% respectively). In the double mutant line *tfiis/iws1* IR is the only AS event that is represented whereas in *tfiis* the three most relevant AS events including IR all represent 33.33% of all AS events (Figure 9 A,C).

The second most prominent event of AS belongs to the class of 5'SS. When applying less stringent filters to identify genes with altered alternative splicing events ( $p < 0.01$ ,  $p < 0.05$ ) IR and 5'SS are the most prominent AS events in all analysed mutant plant lines (Figure 9).

The proportions of the other AS events are smaller indicating that 3'SS, exon skipping and mutually exclusive exon play minor roles in AS compared to IR and 5'SS.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



**Figure 9. Alternative splicing events differentially affected in *tef* mutants compared to *Col-0*.** Alternative splicing events detected by RNAseq were compared between *Col-0* and the mutant lines *tflls* (A), *iws1* (B), *spt4R3* (C) and the double mutant line *tflls/iws1* (D). Percentages of genes differentially alternatively spliced (AS events) in the respective *tef* mutant line compared to *Col-0* are highlighted in yellow (intron retention), grey (alternative 5' splice site), orange (3' alternative splice site), light blue (exon skipping), dark blue mutually exclusive exon. Filters with different stringencies were applied to identify differentially alternatively spliced genes ( $p < 0,005$  high stringency,  $p < 0,05$  low stringency).

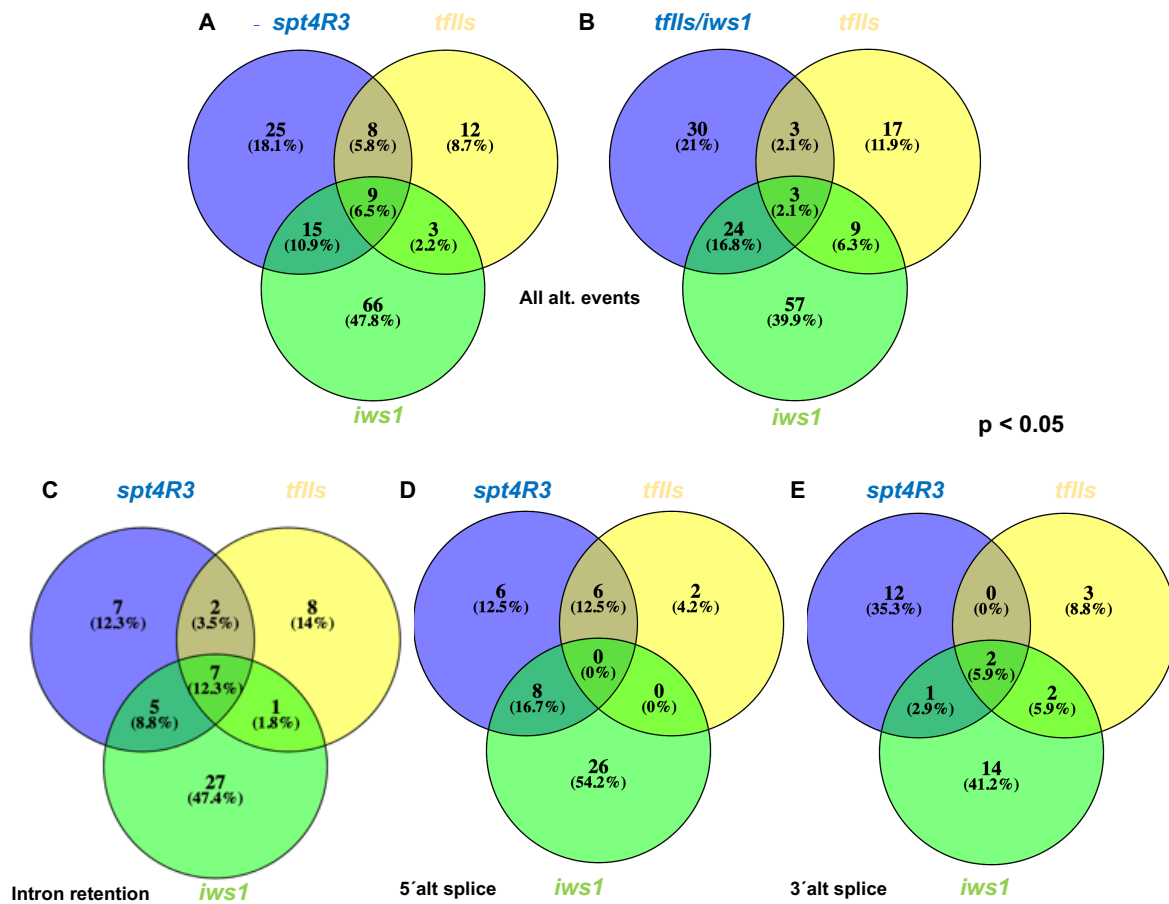
When comparing the genes differentially alternatively spliced in the single mutant lines *iws1*, *tflls* and *spt4R3* it can be observed that the most genes (93) are differentially alternatively spliced in *iws1* and about half of these genes (47.8%) are exclusively altered in this single mutant genotype (Figure 10 A). In *tflls* and *spt4R3* on the other hand less genes (32 and 57 respectively) are differentially alternatively spliced (Figure 10 A).

When analysing whether genes differentially alternatively spliced are altered in more than one of the three single mutant genotypes it can be observed that the different TEFs affect AS of different genes. Of all differentially alternatively spliced genes only 6.5% are affected in all three single mutant genotypes *tflls*, *spt4R3* and *iws1* (Figure 10 A). The majority of all differentially alternatively spliced genes are exclusively altered in one single mutant genotype (~75%, Figure 10 A). This pattern can also be detected when analysing specific AS events. Only 12.3% and 5.9% of all differentially affected alternatively spliced genes are affected in all three single mutant lines when looking at the AS events IR and 3' SS respectively (Figure 10 C,E). No overlap of alternatively spliced genes in the three single mutant lines can be detected when looking at the AS event of 5' SS (Figure 10 D).



## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

A variety of genes are exclusively differentially alternatively spliced in the double mutant *tflls/iws1* but not in the corresponding single mutant lines *tflls* and *iws1* (Figure 10 B).



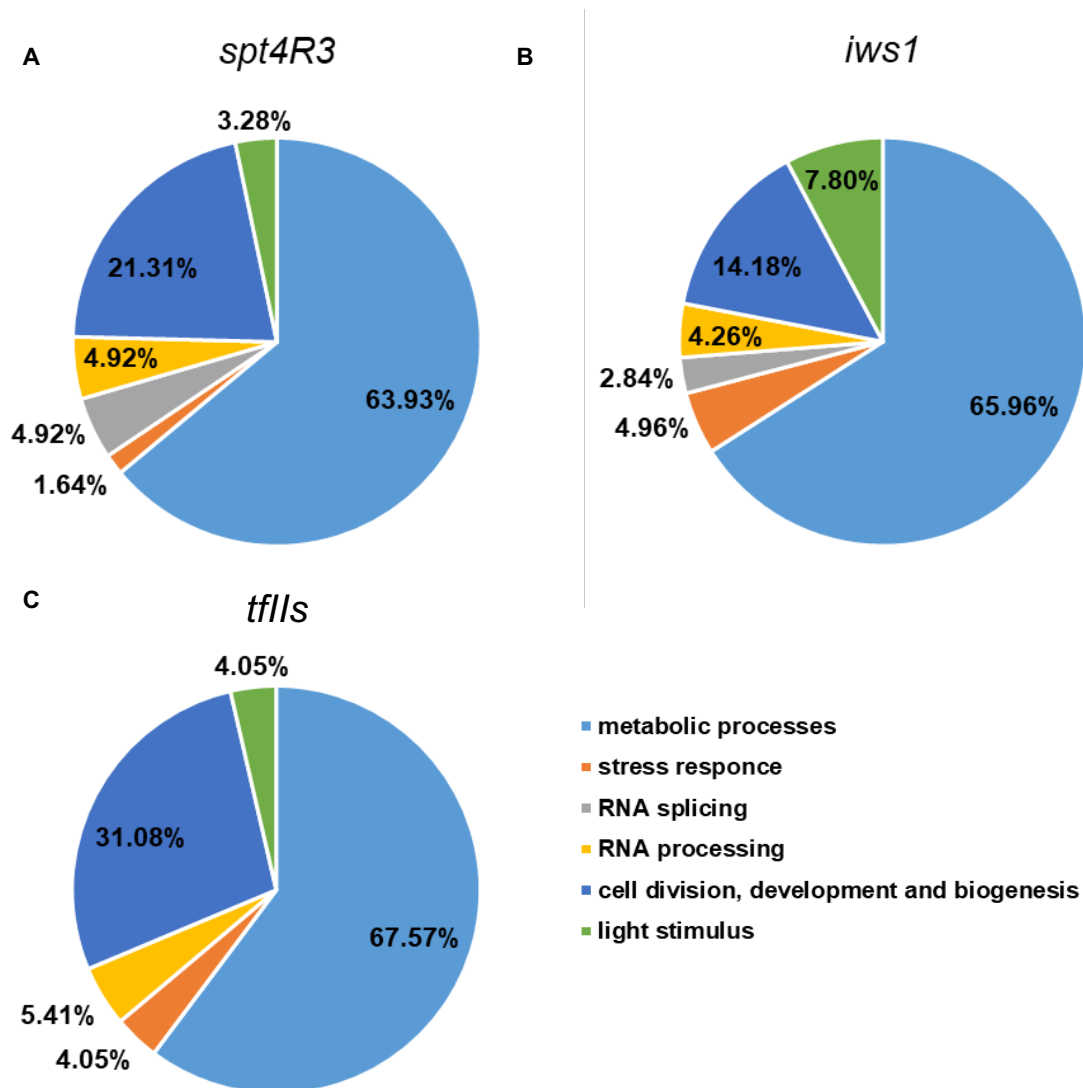
**Figure 10.** Comparison of differentially alternatively spliced genes in *tef* mutants. Overlap of differentially alternatively spliced genes in *tflls*, *iws1*, *spt4R3* (A) and *tflls/iws1*, *tflls*, *iws1* (E). B, C and D) overlap of identified genes representing the AS events of intron retention (B), 5' SS (C) and 3' SS (D) in the single mutants *tflls*, *iws1* and *spt4R3*.

## 2.3 Gene Ontology analysis of genes affected in AS

Gene Ontology (GO) terms are widely applied to understand the biological significance of genome wide analysis. The genes differentially alternatively spliced in the *tef* mutant lines were analysed for GO category enrichment regarding their biological function using *Plaza* (3.0 Dicots Workbench) according to the false discovery rate (FDR < 0.05). In studies on *Arabidopsis* where the effect of different nutrient conditions was tested, differentially expressed genes were involved in mechanisms depending on the type of stress (Nishida et al., 2017). In light stress experiments a strong enrichment in GO categories associated with RNA processing and AS was observed among genes whose AS patterns were affected by a brief light pulse (Mancini et al., 2016). Analysis of the functional categories of genes associated with light-regulated AS showed an overrepresentation of the terms "RNA" and "metabolism" for blue light and "RNA" for white light (Hartmann et al., 2016).

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

The analysis in this work revealed enriched GO terms which are shared in mutant lines vs Col-0. Most of the candidate genes affected in AS can be associated with metabolic processes with 63.93% of genes in *spt4R3* (Figure 11 A, Supplementary table 2), 65.96% in *iws1* (Figure 11 B, Supplementary table 8) and 67.57% in *tflIs* representing this GO term (Figure 11 C, Supplementary table 5). Apart from metabolic processes also the GO terms cell division, development and biogenesis, RNA processing, stress response and light stimulus are represented by candidate genes in all mutant genotypes.



**Figure 11. Gene ontology enrichment analysis of genes differentially alternatively spliced in the different *tef* single mutants.** GO enrichment was evaluated at the level of biological processes in the mutant genotypes *spt4R3* (A), *iws1* (B) and *tflIs* (C) compared to Col-0. Proportions of genes involved in metabolic processes are highlighted in light blue, development and biogenesis in dark blue, RNA processing in yellow, light stimulus in green, stress response in orange and splicing in grey.

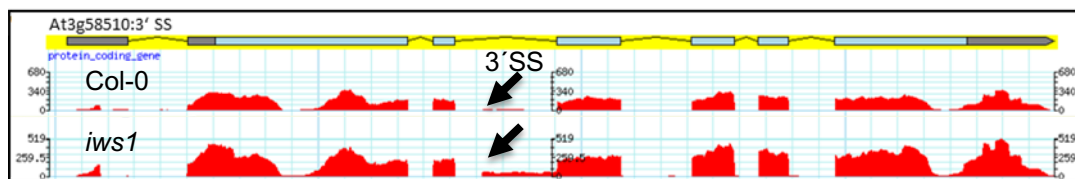
Most of these candidates belong to the group of genes involved in cell division, development and biogenesis (21.31% in *spt4R3*, 14.18% in *iws1* and 31.08% in *tflIs*, Supplementary table 3, 6, 9,

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

Figure 13) and smaller subsets of genes are involved in RNA processing, stress response and light stimulus (Supplementary table 4, 7, 10 Figure 11). The GO term RNA splicing was only enriched among genes differentially alternatively spliced in the mutants *spt4R3* and *iws1* (4.92% and 2.84% respectively, Figure 11, Supplementary table 4, 10).

### 2.4 Experimental validation of differentially alternatively spliced genes

From the list of candidate genes that were differentially alternatively spliced in the *tef* mutant lines according to the RNAseq analysis ( $p < 0.05$ ), 10 candidates were selected and experimentally validated (Table 1). To screen for candidates to be experimentally validated, RNAseq reads of differentially alternatively spliced genes were analysed and coverage plot reads were compared in the respective *tef* mutant(s) and Col-0. An example of different cover reads between the *tef* mutant *iws1* and Col-0 is shown in Figure 12 for the gene At3g58510 (3'SS). The cover reads of the other nine genes that were used to validate the splicing events are shown in Supplementary Figure 1.



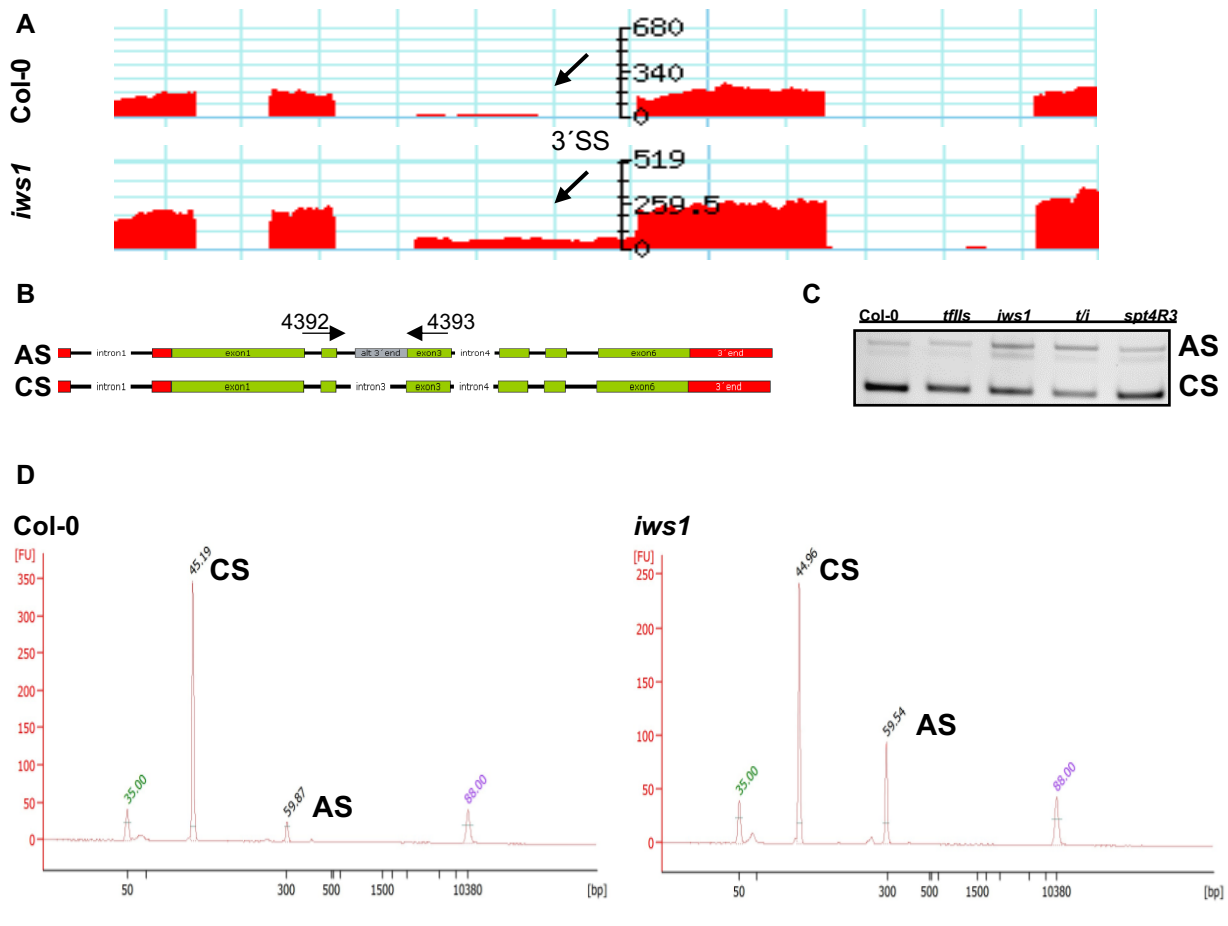
**Figure 12.** RNAseq cover reads of the differentially alternatively spliced gene At3g58510. Gene model of the differentially alternatively spliced gene At3g58510 and corresponding plotted RNAseq reads in Col-0 and the mutant *iws1*. Coverage plot reads within the AS regions are shown in red. A significant difference in the 3'SS event for At3g58510 (second intron) between Col-0 and *iws1* can be detected.

For the experimental validation of differentially alternatively spliced events in the *tef* mutants genes were selected that showed a clear difference in reads between Col-0 and the respective *tef* mutant(s) in the region of the differentially alternatively spliced event. Selected genes also covered different functional categories including genes encoding splicing factors, RNA binding proteins, RNA helicases, dehydrogenases, stress proteins, transmembrane proteins, receptor proteins and kinase proteins. The workflow for the experimental validation of the 10 different candidate genes is shown in Figure 13. Candidates were selected by comparing the mapped RNAseq reads at the region of the differentially alternatively spliced event between Col-0 and *tef* mutant (Figure 13 A). In Figure 15 gene At3g58510 is shown as an example that is differentially alternatively spliced in *iws1* with a higher ratio of an alternatively spliced 3'SS isoform (AS) neighbouring exon 3 to a constitutively spliced isoform (CS) lacking this 3'SS (Figure 13 A,B). To visualize differences among genotypes, RT-PCR was performed using primer pairs amplifying both major splicing isoforms (Figure 13 B,C). To determine the exact length and sequence of the two major splicing isoforms, PCR products obtained by RT-PCR were additionally sequenced (Figure 13 C, Supplementary Figure 2). To quantify the amount of the two major isoforms capillary electrophoresis using a bioanalyzer was performed (Figure 13 D, Supplementary Table 11). This experimental approach is commonly used

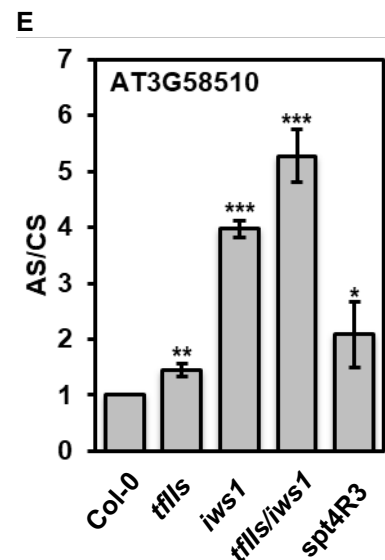
## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

in such experiment (Drechsel et al., 2013; Hartmann et al., 2016; Ruhl et al., 2012). The two main peaks represent the two major splicing isoforms whereby the height of the peak is a measure for the quantity of the respective DNA fragment. This quantification was used to calculate the ratios between the AS isoform and the CS isoform in Col-0 and the respective *tef* mutants (Figure 13 D, Supplementary Figure 3). For the differentially alternatively spliced gene At3g58510 a significantly higher ratio of AS isoform to CS isoform was measured in all *tef* mutants (Figure 13 E).

2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



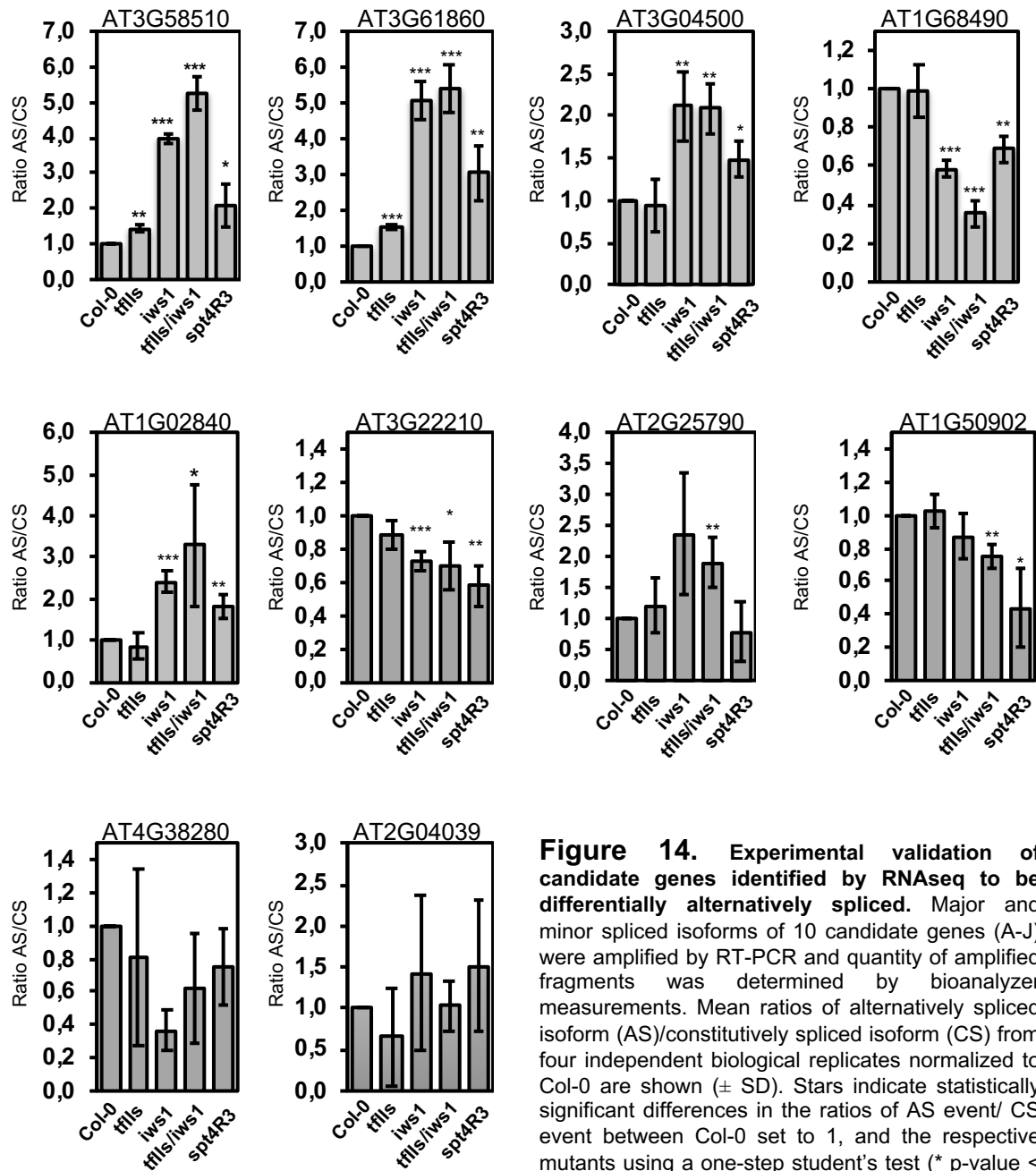
**Figure 13.** The validation of 10 candidate genes differentially alternatively spliced in at least one *tef* mutant according to RNAseq. Candidates were selected by comparing mapped reads at regions alternatively spliced in Col-0 and the *tef* mutants (A). For candidate gene At3g58510 more reads of an alternatively spliced isoform with an alternative 3'SS next to exon 3 (B, AS) were identified in *iws1* compared to Col-0 (A, B). Major and minor spliced isoforms were amplified by RT-PCR (C) and quantified by bioanalyzer measurements (E). Mean ratios of alternatively spliced isoform (AS)/constitutively spliced isoform (CS) quantified by bioanalyzer measurements from four independent biological replicates were calculated and normalized to Col-0 ( $\pm$  SD) with stars indicating statistically significant differences in the ratios of AS event/ CS event between Col-0 and the respective mutants using student's test (\* p-value < 0.05, \*\* p-value < 0.01, \*\*\* p-value < 0.001).



The experimental validation shown in Figure 13 was performed on 10 different candidate genes (Figure 14, Table 1). For all candidate genes it was possible to amplify and quantify the two major splicing isoforms. Within the 10 different candidate genes 18 differentially alternatively spliced events were detected by RNAseq and 14 times the respective candidate gene was also differentially alternatively spliced in the respective *tef* mutant according to the experimental validation (Table 1).

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

For some candidate genes that are differentially alternatively spliced in a specific *tef* mutant according to RNAseq, in parts a clear, but not statistically significant difference in the splicing pattern could be detected in the respective mutant compared to Col-0. For example the ratio of AS/CS event for the candidate gene *At2g25790* was more than 2 times higher in *iws1* compared to Col-0 but not statistically different (Figure 14). In addition to the affected AS events detected by both, RNAseq and experimental validation, 10 times candidate genes showed a statistically significant change of the AS/CS ratio in a specific *tef* mutant only in the experimental validation but not in the RNAseq experiment (Figure 14, Table 1).



## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

RNAseq and experimental validation (RT-PCR, isoforms products sequencing and quantification by bioanalyzer) showed that compared to IWS1 and SPT4 the TEF TFIIIS might have a minor impact on AS in *Arabidopsis*. According to the experimental validation only the candidate genes At3g61860 and At3g58510 are differentially alternatively spliced in *tflls* (Table 1). In *iws1*, *tflls/iws1* and *spt4R3* the splicing patterns in the candidate genes are more severely affected with 6 (*iws1*), 8 (*tflls/iws1*) and 8 (*spt4R3*) candidate genes differentially alternatively spliced (Table 1) indicating that the TEFs IWS1 and SPT4 have a bigger impact on AS.

**Table 1. Candidate genes used for experimental validation.** 10 candidate genes that were differentially alternatively spliced in at least one mutant genotype according to RNAseq and that were used for experimental validation of the RNAseq results by RT-PCR, PCR product sequencing and Bioanalyzer measurement. Stars indicate statistically significant differences between Col-0 and the respective mutants regarding AS/CS ratios based on RNAseq or quantification/validation by Bioanalyzer measurement (\* p-value < 0.05, \*\* p-value < 0.01, \*\*\* p-value < 0.001). Student's t-test was applied to test for significant differences of AS/CS ratios determined by Bioanalyzer measurements.

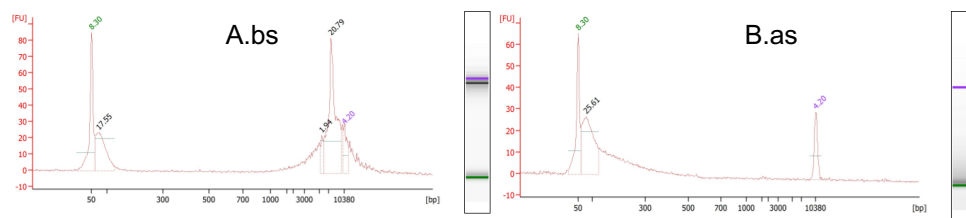
	<i>tflls</i>		<i>iws1</i>		<i>tflls/iws1</i>		<i>spt4R3</i>		
	RNAseq data	validation analysis	RNAseq data	validation analysis	RNAseq data	validation analysis	RNAseq data	validation analysis	AS event identified
At3g61860	**	***	***	***	***	***	***	**	5'SS
At3g04500	ns	ns	**	**	*	**	ns	*	3'SS
At2g25970	ns	ns	*	ns	**	**	ns	*	5'SS
At3g58510	ns	**	***	***	***	***	ns	*	3'SS
At1g02840	ns	ns	ns	***	**	**	*	*	3'SS
At3g22210	*	ns	ns	***	ns	*	***	**	IR
At1g68490	ns	ns	ns	***	***	***	ns	**	5'SS
At1g05090	ns	ns	ns	ns	***	**	ns	*	IR
AT2g04039	ns	ns	ns	ns	ns	ns	*	ns	IR
AT4g38280	**	ns	ns	ns	ns	ns	ns	ns	3'SS

## 2.5 Genome-wide analysis of RNAPII occupancy: from RNAseq to ChIPseq

In the following, genome-wide quantitative profiling was used to investigate the mutual relationships between chromatin, RNAPII occupancy, transcription elongation and the co-transcriptional process of alternative splicing. Therefore Chromatin Immunoprecipitation (ChIP) followed by high-throughput sequencing (ChIPseq) was performed. Since the previous results indicated that the TEFs IWS1 and SPT4 have a bigger impact on AS than TFIIIS (Figure 14, Table 1), ChIPseq analyses were conducted in the mutant lines *iws1* and *spt4R3*. For immunoprecipitation, specific antibodies against the NRPB1 phosphorylated at the CTD position Ser2 (S2P) and Ser5 (S5P) were used.

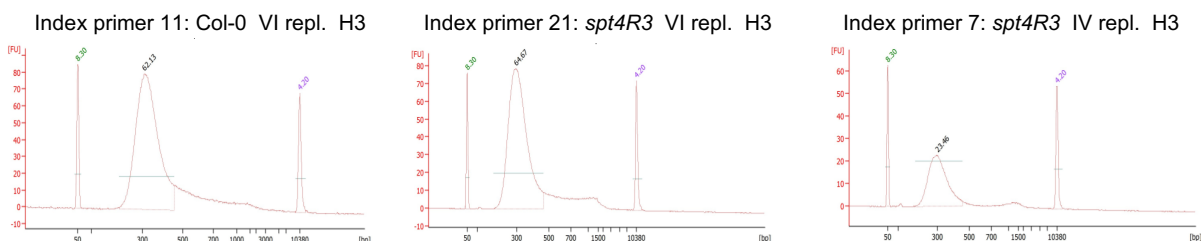
## 2.6 Quality control of ChIP experiments

Before extracted DNAs were used to generate libraries, several quality control steps were applied. Shearing of DNA is a crucial step in ChIP experiments in order to analyse protein-binding regions. The extracted genomic DNA must be sheared into smaller, suitable fragments. DNA fragmentation was achieved mechanically by sonication (Bioruptor® Pico sonication device). Ideal chromatin fragments for sequencing range from 100 to 350 bp. The samples were tested before and after the sonication step with the use of a Bioanalyzer and the shearing of extracted DNA could be confirmed (Figure 15). Before sonication a sharp peak around 10.000 bp is depicted that represents the chromatin state before the mechanic fragmentation (Figure 15 A). After sonication the peak at 10.000 bp is completely gone and a larger peak is present around 100bp (Figure 15 B), demonstrating that the DNA was efficiently fragmented.



**Figure 15. Bioanalyzer analysis of a chromatin sample before and after sonication.** A) Before sonication (bs) a peak around 10.000 bp represents the chromatin DNA at the expected size before fragmentation. The number above the peaks indicate the concentration in ng/ul. B) after sonication (as) the peak at 10.000 bp is clearly reduced and one larger peak around 100 bp is present showing that the DNA was efficiently fragmented.

Before the DNA libraries were prepared, additionally the DNA concentrations after immunoprecipitation were measured using a Qubit device (Supplementary Table 12) and the size distributions of adapter-ligated fragments (200-300 bp) were determined by Bioanalyzer measurements (Figure 16, Supplementary Figure 6).



**Figure 16. Size distribution of adapter-ligated fragments in representative samples determined by Bioanalyzer measurement.** The peaks around 300bp represent the distribution of the samples after adaptor ligation and PCR amplification. The numbers above the peaks indicate the concentrations in ng/ul. In green and violet respectively the ng/ul of upper and lower mark.

The generated sequencing reads were aligned to the *Arabidopsis* reference genome (TAIR10) in our lab by Simon Obermayer. The number of mapped reads are listed in Supplementary Figure 4



(~14 mio raw reads per sample). In all ChIPseq experiments performed in this study reproducibility between biological replicates was assessed using the principal component analysis (PCA) and Spearman correlation of individual replicates (Supplementary figure 5). Using Ser2 and Ser5 CTD specific antibodies for immunoprecipitation one replicate each did not cluster. These replicates were consequently not included in the analysis.

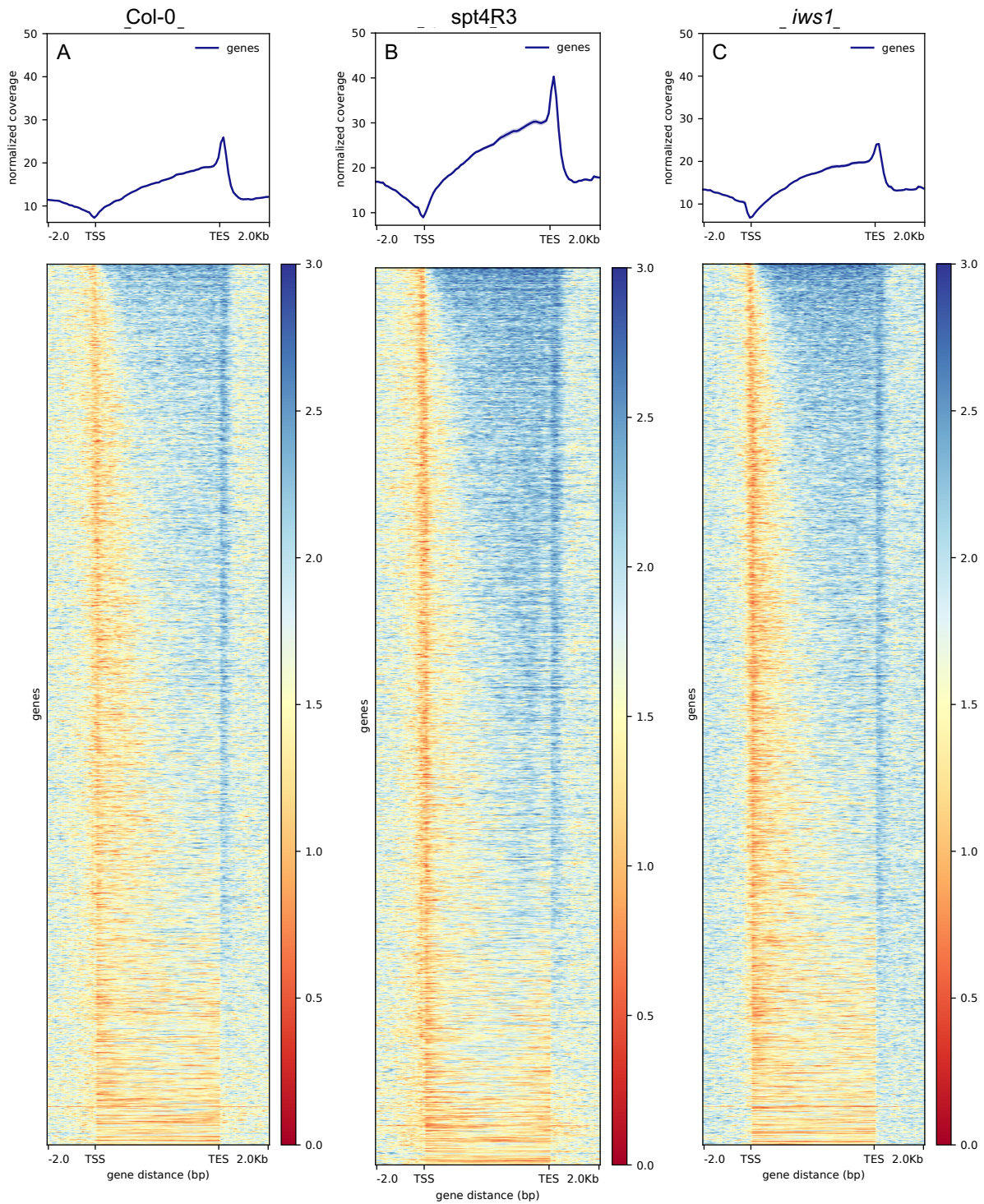
## 2.7 RNAPII occupancy in *tef* mutants

Several studies in plants showed that RNAPII elongation influences splicing (Jabre et al., 2019; Zhu et al., 2018). To expand the knowledge about this correlation, genome wide ChIPseq experiments analysing the RNAPII distribution in the *tef* mutant lines *iws1* and *spt4R3* was performed. *IWS1* and *SPT4* were chosen since the previous results indicated that these TEFs have a bigger impact on AS than TFIIS (Figure 14, Table 1). RNAPII Ser2P and RNAPII Ser5P occupancy was determined over all TAIR10 annotated genes and was depicted from 2 kbp upstream the transcription start site (TSS) to 2 kbp downstream the transcription end site (TES). TSS and TES encompass a region of 5kbp. In this study only one biological replicate could be analysed after ChIPseq and thus the following results might lack reproducibility. The genes were sorted and colour scaled according to length from longer genes to shorter genes. The standard error is almost zero (Figure 17) showing the good quality of the gene distribution.

RNAPII Ser2P increases over gene bodies and shows an accumulation at the transcription end site (TES) (Figure 17). A similar genome-wide distribution of Ser2P RNAPII profiles was also observed in other publications (Antosz et al., 2020; Hetzel et al., 2016; Zhu et al., 2018) and in a previous study performed in our work group (PhD Antosz, 2019). In Col-0 and *iws1* mutant line a similar RNAPII Ser2P distribution can be observed albeit in *iws1* a slightly higher occupancy over the gene body compared to Col-0 can be detected (Figure 17 A,C). In the *spt4R3* mutant line on the other hand Ser2P density was enriched over the gene body and a higher peak at the TES is depicted (Figure 17 B).

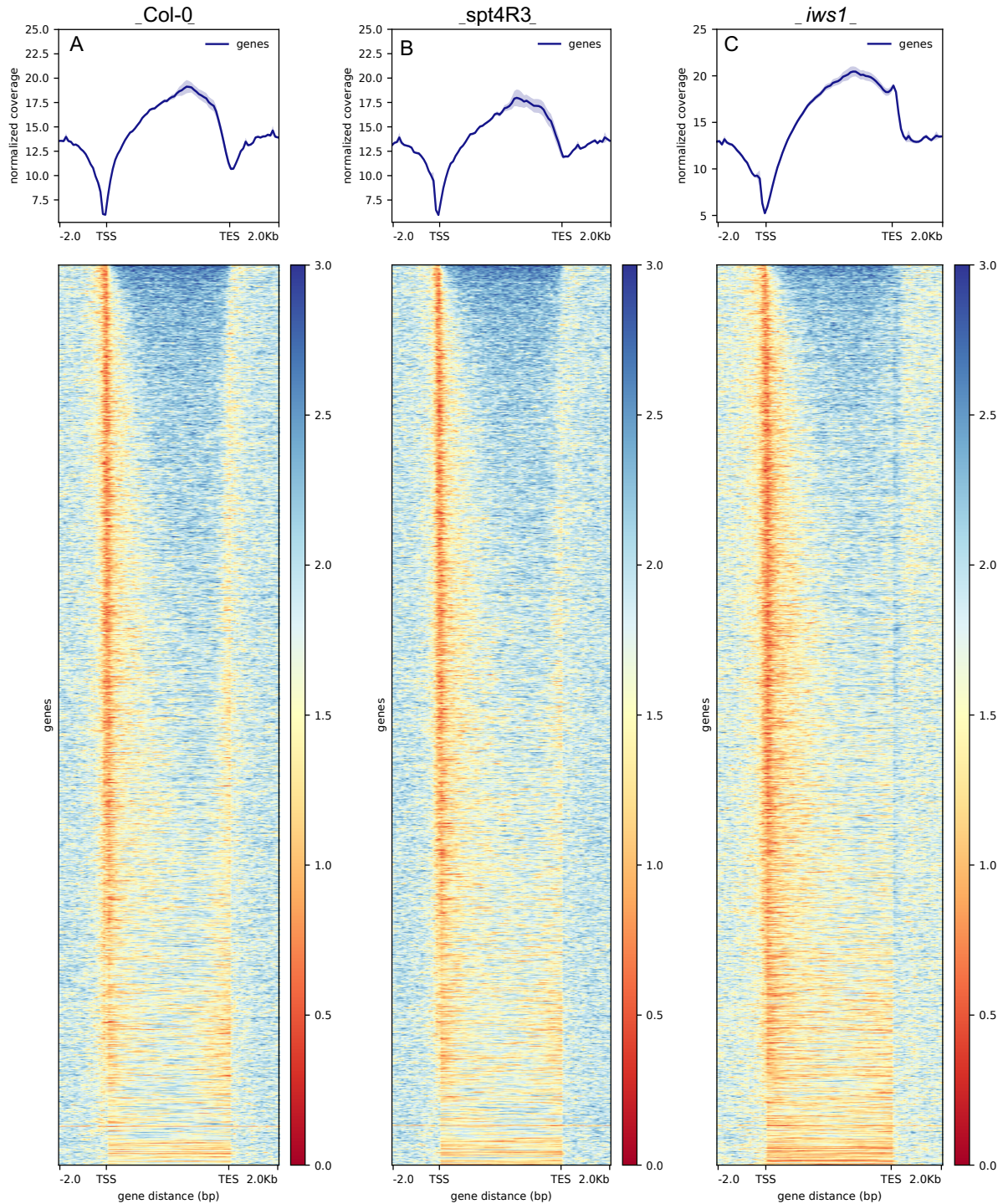
In contrast to the Ser2P profile, in Col-0 Ser5P RNAPII densities are enriched immediately downstream the TSS and the occupancies along the gene bodies get reduced before the TES (Figure 18). The shadows on the peaks represent the standard error. In the *iws1* mutant line the ChIPseq profile indicates a higher Ser5P accumulation along the gene bodies and an defined sharp peak at the TES, whereas the Ser5P RNAPII profile in the *spt4R3* mutant line is not different compared to Col-0.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



**Figure 17. RNAPII Ser2P occupancy profile over all TAIR10 annotated genes determined by ChIP-seq.** Comparison of RNAPII Ser2P accumulation between Col-0 (A), *spt4R3* (B) and *iws1* (C). RNAPII occupancy is displayed in the range between -2 kb upstream of the TSS and 2kb downstream of the TES. TSS and TES encompass a region of 5kbp. All genes from TAIR 10 are represented and sorted according to region length from longer to shorter. Genes are coloured scaled. ChIP tracks are normalized according input libraries. The shadow on the profile distribution represents the standard error. Figures were created by Simon Obermeyer.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



**Figure 18.** RNAPII Ser5P occupancy profile over all TAIR10 annotated genes determined by ChIP-seq. Comparison of RNAPII Ser5P accumulation between Col-0 (A), *spt4R3* (B) and *iws1* (C). RNAPII occupancy is displayed in the range between -2 kb upstream of the TSS and 2kb downstream of the TES. Region body length 5kb. All genes from TAIR 10 are represented and sorted according to region length from longer to shorter. Genes are coloured scaled. The shadow on the profile distribution represents the standard error. Figures were created by Simon Obermeyer.

## **2.8 The phosphorylation status of RNAPII is correlated with pausing at intron/exon boundaries**

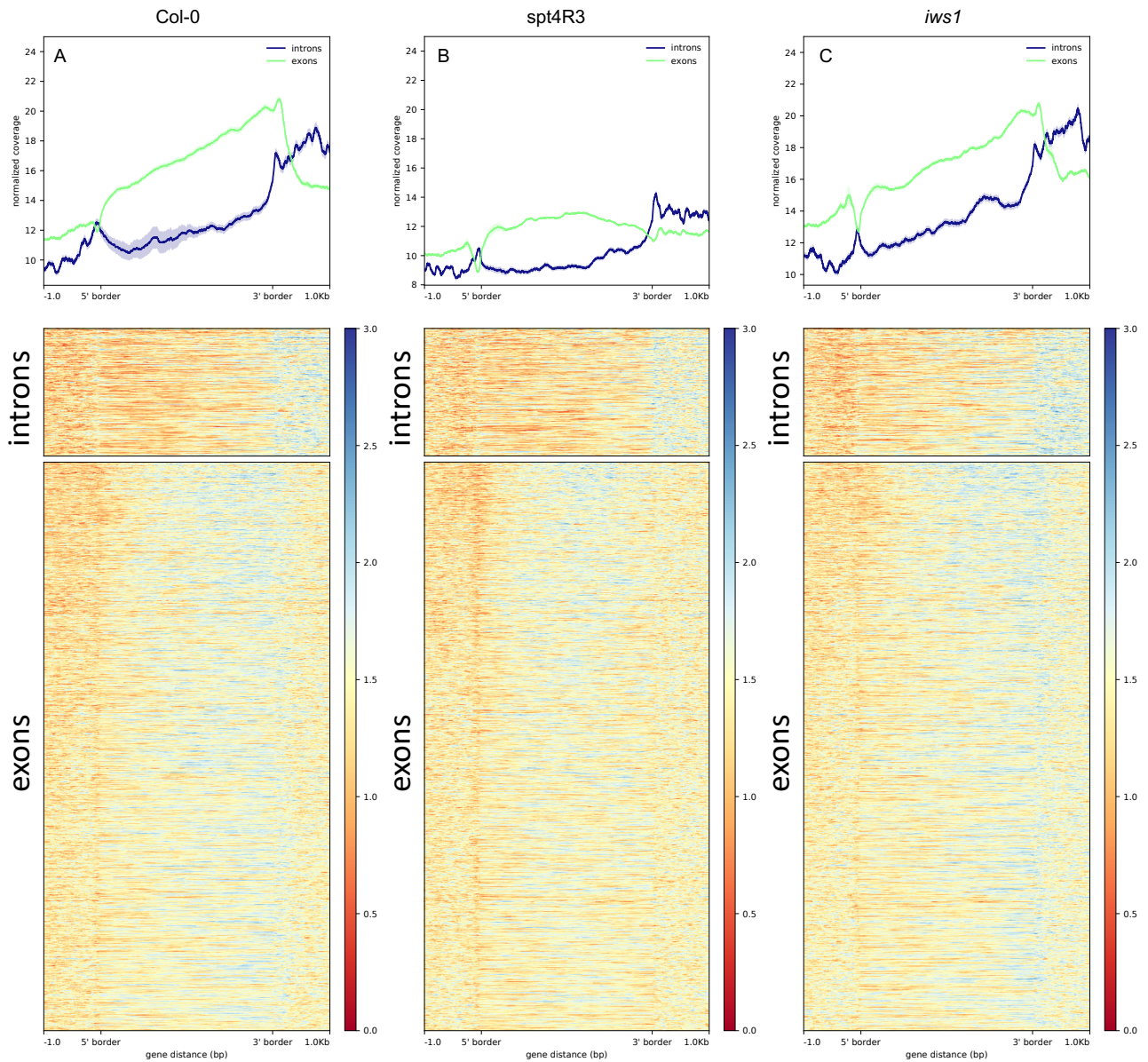
In human cells it was demonstrated that pausing of RNAPII at 3' and 5' splice sites is accompanied with increased Ser5 CTD phosphorylation whereas the release from splice site associated pauses is accompanied with an increase in Ser2 CTD phosphorylation (Saldi et al., 2016).

ChIPseq also revealed some differences of RNAPII Ser2P and Ser5P occupancies at the 5'SS and 3'SS of intron-exon boundaries. All introns and exons were isolated from the TAIR10 genome annotation and additionally filtered for introns and exons that are larger than 500bp. In the following all introns and exons were scaled to the same length (5' border to 3' border) and the RNAPII Ser2P and Ser5P distributions over introns and exons were analysed (Figure 19, 20).

Both, Ser2P and Ser5P modified RNAPIIs show a peak at 5' and 3' borders of introns demonstrating that RNAPII slows down at intron/exon boundaries (Figure 19, 20). Ser2P RNAPII distributions are reduced at 5' borders of exons and densities gradually increase towards the 3'end of the exon (Figure 19). Ser5P RNAPII distributions are also reduced at 5' borders of exons, Ser5P RNAPII levels increase towards the middle of the exons before the density gets reduced towards the 3'end of the exon (Figure 19).

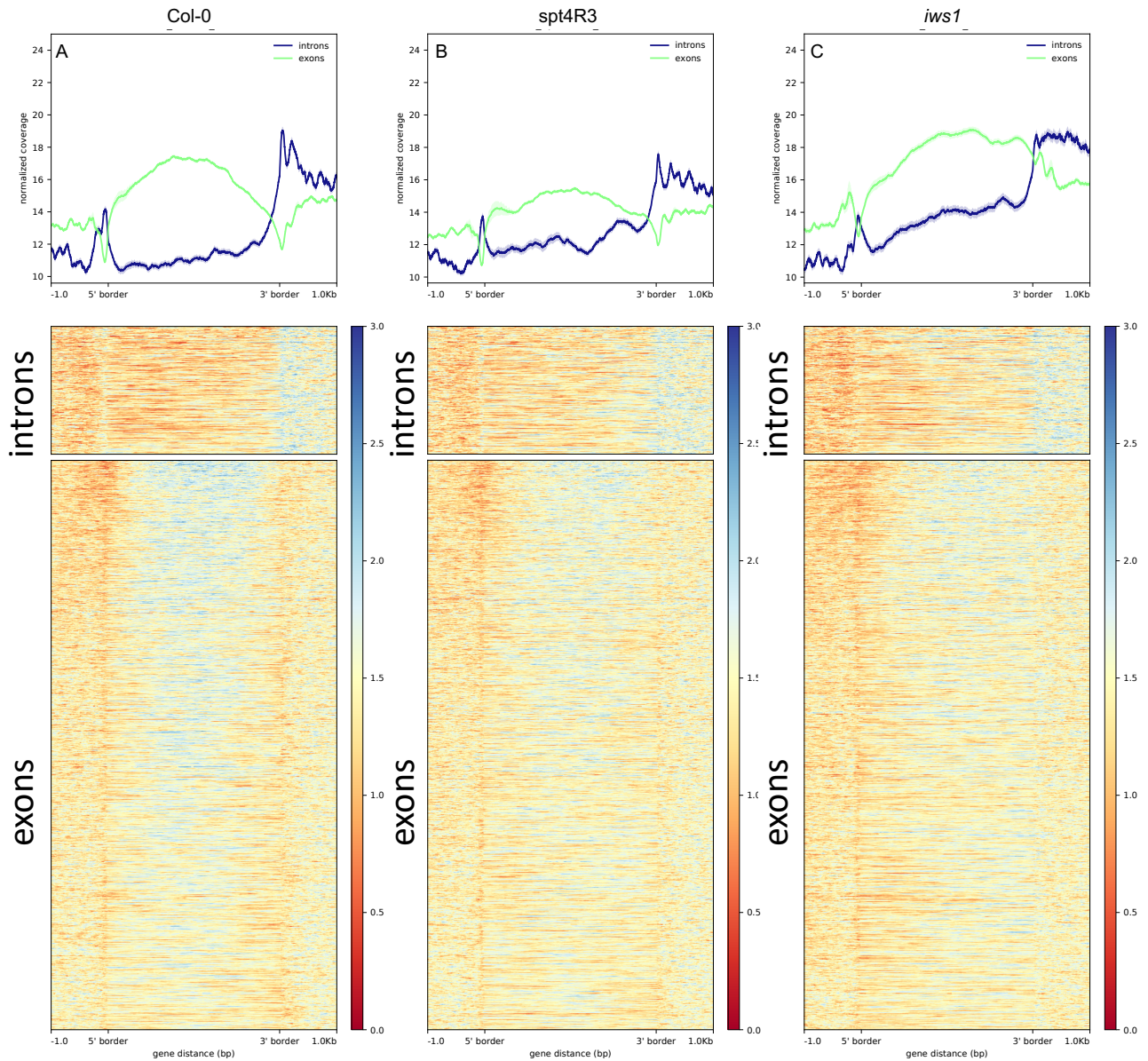
In general the Ser5P modified RNAPII shows a stronger accumulation at intron/exon boundaries indicating that the RNAPII that gets slower at splice sites is predominately phosphorylated at Ser5 positions (Figure 19, 20). The shadows on the profile distributions represent the standard error.

## 2. Results: The effect of transcription elongation factors on alternative splicing in Arabidopsis thaliana



**Figure 19.** Mean RNAPII Ser2P distribution over exons and introns measured by ChIP-seq. Density line plots and heat map reads of A) Col-0, B) *spt4R3* and C) *iws1*. Introns and exons were isolated from the TAIR10 genome annotation and filtered for introns and exons that are larger than 500bp. Introns (blue lines) and exons (green lines) were scaled to the same length (5' border to 3' border). Genes are coloured scaled. The shadow on the profile distribution represents the standard error. Figures were created by Simon Obermeyer.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*



**Figure 20** Mean RNAPII Ser5P distribution over exons and introns measured by ChIP-seq. Density line plots and heat map reads of A) Col-0, B) *spt4R3* and C) *iws1*. Introns and exons were isolated from the TAIR10 genome annotation and filtered for introns and exons that are larger than 500bp. Introns (blue lines) and exons (green lines) were scaled to the same length (5' border to 3' border). The shadow on the profile distribution represents the standard error. Figures were created by Simon Obermeyer.

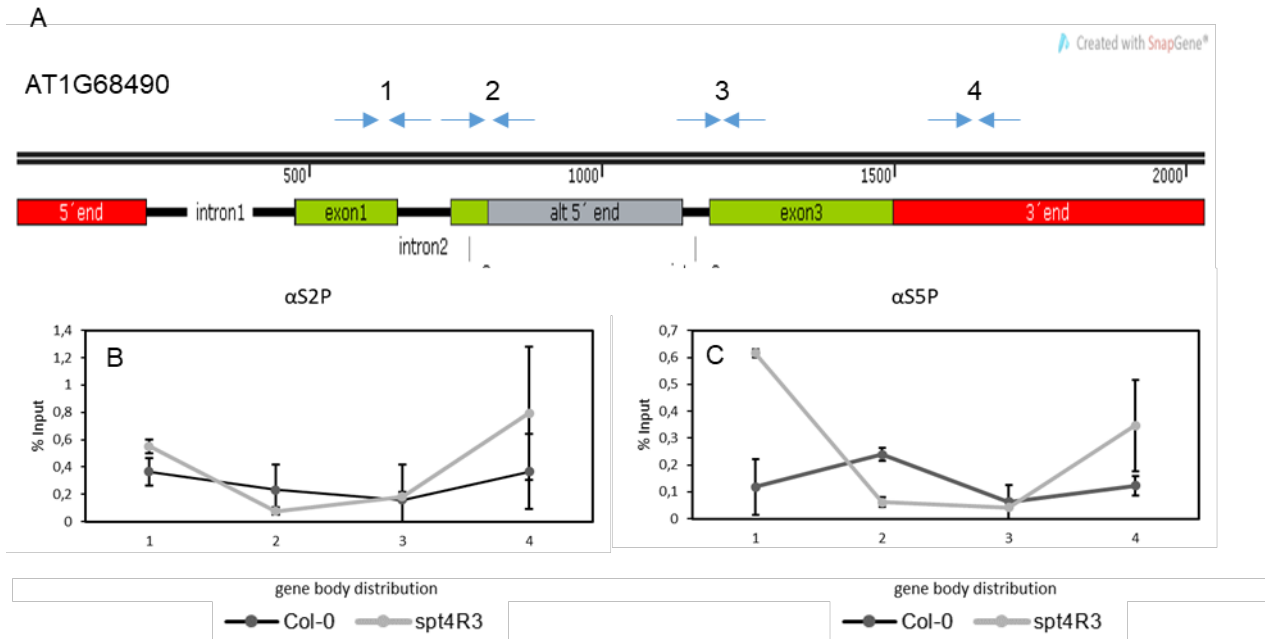
## 2.9 Validation of RNAPII ChIPseq results by ChIP-qPCR

The coupling between phosphorylation status of the RNAPII, transcription elongation and pre-mRNA splicing could be demonstrated in animals and plants (Nojima et al., 2015; Zhu et al., 2018). ChIPseq performed in this study could expand the knowledge about these correlations in *Arabidopsis*. In the following it was analysed if there is a correlation between the global RNAPII distribution pattern and the RNAPII occupancy over a representative gene which is differentially alternatively spliced in the *tef* mutant lines *iws1* and *spt4R3*. The alternative spliced gene At1g68490 displays an alternative 5'SS between exon 2 and exon 3 (Figure 21 A). In the course of the experimental validation of the RNAseq data it could be shown that in the *tef* mutant lines *iws1* and *spt4R3* this alternative spliced version is significantly altered (Figure 14, Table 1). By qPCR performed on one single gene one cannot draw conclusions on the global RNAPII distribution profile. The intention of this experiment was to cross-correlate the RNAseq results with the ChIPseq and thus this experiment served as a quality control.

RNAPII occupancies along At1g68490 were further analysed at individual loci by qPCR after ChIP. Four primer pairs were used to amplify representative fragments along the gene body. Primer pair 1 binds upstream of the AS event in exon 1. Primer pair 2 binds in exon 2 and the alternative 5' end marking the constitutive 5'SS. Primer pair 3 anneals to intron 3 and exon3 marking the 3'SS and primer pair 4 amplifies a region in the 3'UTR region of the gene (Figure 21 A). As a negative control all ChIP experiments performed in this study were also carried out without the use of specific antibodies. QPCR signals in mock treated samples were reduced to background levels in all experiments (data not shown).

The results obtained could confirm the genome-wide distribution observed by ChIP-seq and the RNAseq data. In the *spt4R3* mutant line a higher Ser2P and Ser5P density at exon 1 (primer pair 1) could be detected (Figure 21 A,B). The difference in the RNAPII occupancy at this region between Col-0 and *spt4R3* is more pronounced regarding the Ser5P modified polymerase. Around the constitutive 5'SS (primer pair 2) on the other hand RNAPII Ser2P and Ser5P distribution is to some extent enriched in Col-0. No differences between RNAPII Ser2P and Ser5P occupancies in Col-0 and *spt4R3* could be observed around the 3'SS (primer pair 3). Towards the 3' end of the gene (primer pair 4) a distinct higher occupancy of Ser2P and Ser5P modified RNAPII was determined by qPCR in the *spt4R3* mutant line.

## 2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

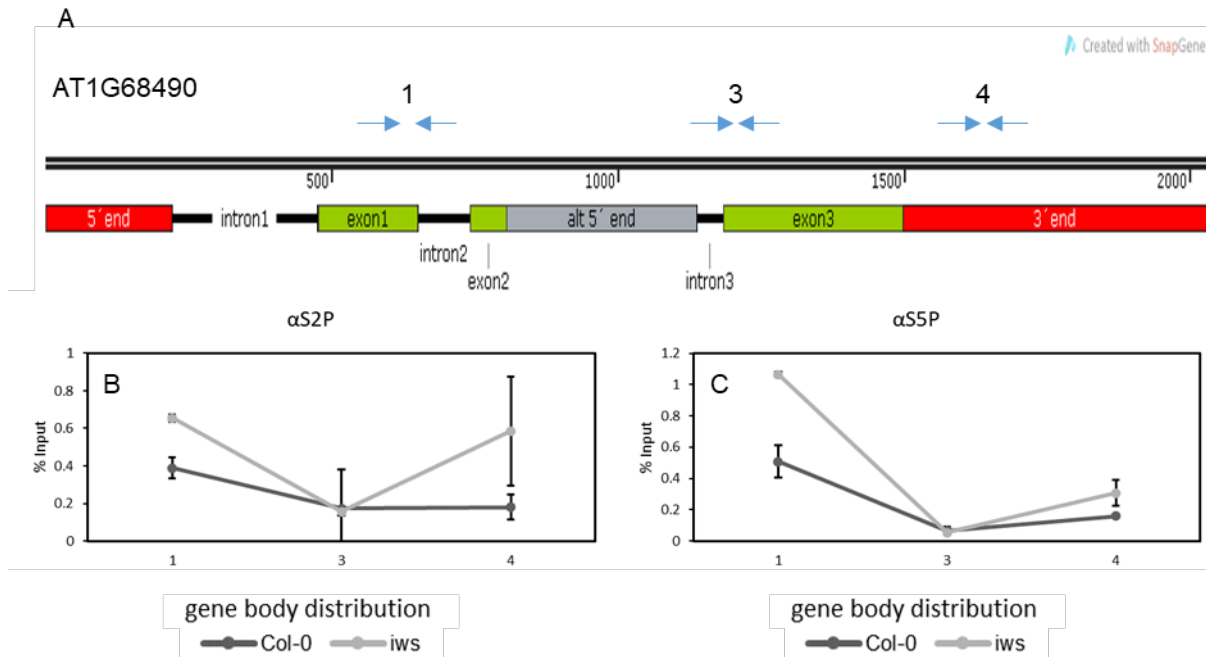


**Figure 21. ChIP-qPCR validation experiment comparing the RNAPII distribution over At1g68490 in Col-0 and *spt4R3*.** A) Schematic representation of At1g68490 with the AS event (5'SS) depicted in grey. Green boxes represent exons and red box 5'UTR and 3'UTR. The arrows above indicate the relative position of primer pairs (1-4) used for qPCR. B, C) qPCR analysis after chip using S2P (B) and S5P (C) specific antibodies. Results are displayed as percentages of the input as determined by qPCR (Y-axis). Error bars represent  $\pm$  SD of at least two biological and two technical replicates. X-axis displays primer pairs (A) used for qPCR.

Considering that in this experiment differences in the distribution were mainly observed around the region amplified with the primer pair combinations 1, 3 and 4, it was decided to continue with this combinations in the follow validation analysis. A similar RNAPII distribution pattern with an enrichment compared to Col-0 at exon 1 (Figure 22 A, primer pair 1) and at the 3'UTR (primer pair 4) can also be observed in *ivs1* (Figure 22 B,C). Additionally also in this mutant line no changes in RNAPII occupancy (primer pair 3) around the 3'SS are detected by qPCR (Figure 22 B,C).



## 2. Results: The effect of transcription elongation factors on alternative splicing in Arabidopsis thaliana













**Figure 22. ChIP-qPCR validation experiment comparing the RNAPII distribution over At1g68490 in Col-0 and *iws1*.** A) Schematic representation of At1g68490 with the AS event (5' SS) depicted in grey. Green boxes represent exons and red box 5' UTR and 3' UTR. The arrows above indicate the relative position of primer pairs (1,3,4) used for qPCR. L,M) qPCR analysis after chip using S2P (L) and S5P (M) specific antibodies. Results are displayed as percentages of the input as determined by qPCR (Y-axis). Error bars represent  $\pm$  SD of at least two biological and two technical replicates. X-axis displays primer pairs (A) used for qPCR.

When analysing the single gene ChIPseq tracks from the gene At1g68490 in the *spt4R3* and *iws1* mutant lines (Supplementary Figure 7) a relative enrichment of Ser2P and Ser5P RNAPII occupancy along the 5' and 3' regions of the gene is also indicated. ChIPseq analysis gives a better overview on the genome-wide scale whereas data get very noisy on single gene scale. For that reason the final conclusion on this work was made on the genome-wide scale rather than on single genes.











Additionally a motif search analysis after RNAseq using the HOMER software was performed to identify putative *cis*-regulatory sequence motifs that are enriched in the differentially alternatively spliced genes around an alternative 5' and 3' SS. The software compares two sets of sequences and tries to identify regulatory elements that are specifically enriched in one set relative to the other. Enriched motifs were next compared with a library of known motifs in humans and sorted based on p-values. Several motifs were detected to be enriched around an alternative 3' and 5' SS indicating that also sequence elements may influence the splicing decisions. The 10 best matched motifs that are enriched at an alternative 3' end (Figure 23 A) and at an alternative 5' end (Figure 23 B) are listed below and are based on highly significant p-values. The highly significance points out that the sequences found are potential crucial motives in the splicing recognition process.

2. Results: The effect of transcription elongation factors on alternative splicing in *Arabidopsis thaliana*

A. 3' SS

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)
1		1e-1308	-3.013e+03	50.07%	17.48%	609.6bp (503.7bp)
2		1e-1238	-2.852e+03	54.87%	21.85%	604.0bp (561.5bp)
3		1e-1207	-2.781e+03	64.50%	30.43%	625.6bp (519.1bp)
4		1e-1080	-2.488e+03	40.15%	12.89%	605.3bp (541.5bp)
5		1e-1070	-2.464e+03	49.77%	19.82%	704.2bp (526.9bp)
6		1e-922	-2.125e+03	66.13%	36.05%	654.0bp (520.1bp)
7		1e-888	-2.045e+03	39.52%	14.44%	771.7bp (534.0bp)
8		1e-808	-1.861e+03	67.09%	38.90%	718.3bp (508.3bp)
9		1e-740	-1.705e+03	61.94%	35.04%	601.5bp (536.8bp)
10		1e-701	-1.616e+03	69.63%	43.41%	605.6bp (520.9bp)

B. 5' SS

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)
1		1e-357	-8.242e+02	65.45%	24.96%	962.7bp (923.8bp)
2		1e-331	-7.631e+02	68.37%	28.91%	1120.1bp (861.9bp)
3		1e-307	-7.082e+02	35.54%	7.00%	817.1bp (758.7bp)
4		1e-259	-5.973e+02	47.87%	16.28%	1534.1bp (903.6bp)
5		1e-230	-5.317e+02	72.17%	38.92%	1153.2bp (940.8bp)
6		1e-226	-5.215e+02	39.91%	12.47%	1172.0bp (815.9bp)
7		1e-162	-3.740e+02	65.48%	37.52%	1151.1bp (918.6bp)
8		1e-158	-3.644e+02	48.54%	22.69%	1147.2bp (929.8bp)
9		1e-149	-3.431e+02	45.19%	20.67%	1428.6bp (933.5bp)
10		1e-142	-3.292e+02	59.81%	33.76%	1330.5bp (819.5bp)

**Figure 23. Motifs enriched around alternatively 5' and 3' splice sites.** Motifs were specifically enriched in genes with an alternatively spliced 3' splice site (A) and 5' splice site (B) that were differentially alternatively spliced in the *tef* mutant lines. The motives are listed according to P-value. Analyses were performed by Simon Obermayer using the Software HOMER.

## 3. Results: The chromatin status and alternative splicing

### 3.1 The *Arabidopsis* H3K36me3 – MRG adaptor complex

In mammalian cells MRG proteins recruit the splicing factor PTB. This recruitment is initiated by an interaction of the MRG15 reader protein with the histone mark H3K36me3. H3K36me3 is deposited by the methyltransferase SETD2 whereas SETD2 itself is recruited to the chromatin by the TEF IWS1. Thus the regulation of splicing by SETD2, MRG15 and PTB is directly related to IWS1 (Lee and Blenis, 2014).

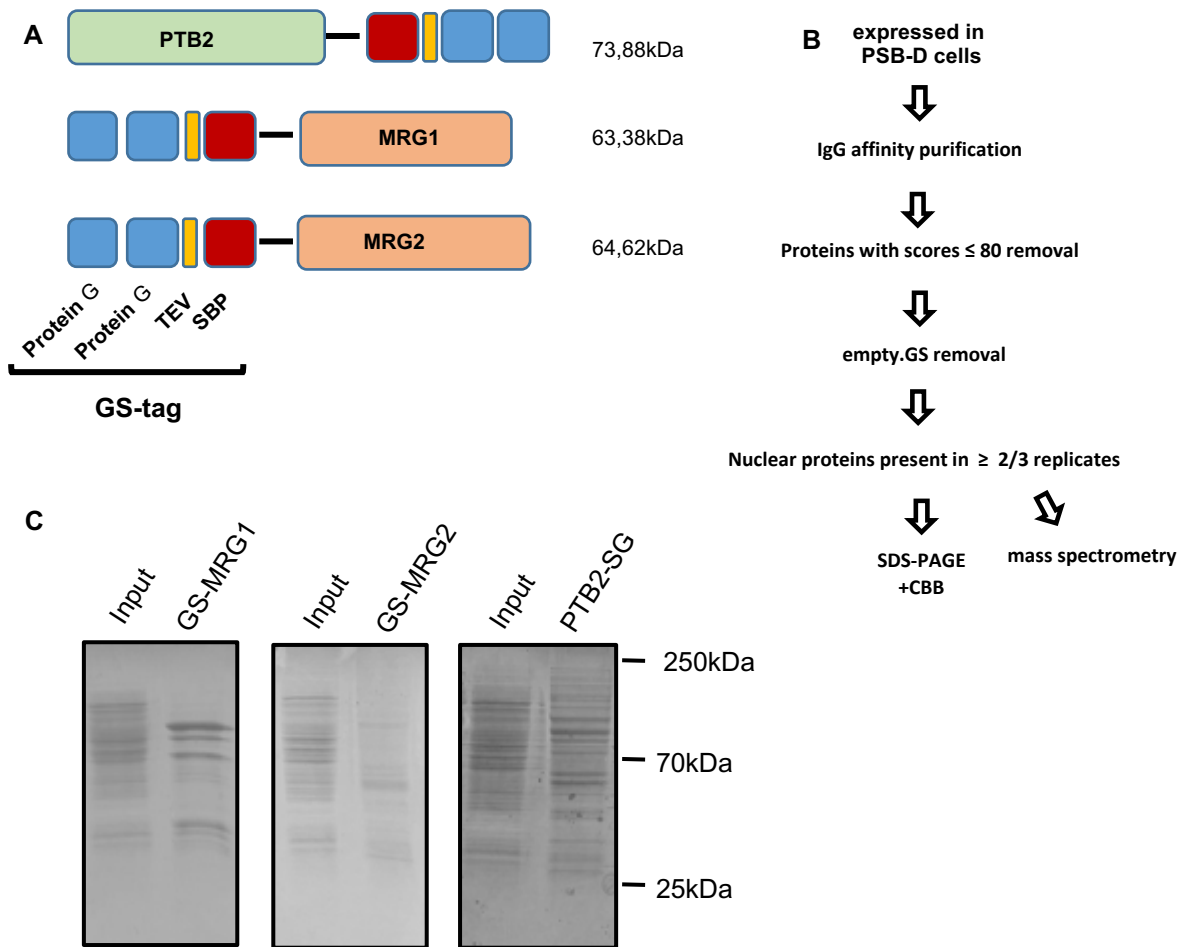
Two orthologues (MRG1 and MRG2) of the human MRG15 protein are described in *Arabidopsis*. They regulate the photoperiodic flowering time by binding directly to FLOWERING LOCUS T (FT) and elevate its expression in an H3K36me3-dependent manner (Bu et al., 2014; Xu et al., 2014).

### 3.2 Splicing factors and the methyltransferase SDG7 co-purify with MRG proteins

To unravel putative interaction partners of MRG1, MRG2 and PTB2 in *Arabidopsis*, affinity purification coupled to mass spectrometry (AP-MS) was performed using protein extracts from *Arabidopsis* cell suspension cultures. For purification, the coding sequences of *MRG1*, *MRG2* and *PTB2* were fused N-terminally or C-terminally to a GS-Tag (Figure 26 A). The transgenes under control of a CaMV 35S promoter were used to transform *Arabidopsis* cell suspension cultures (PSB-D) by co-cultivation with transgenic *Agrobacterium* (Van Leene et al., 2011, 2015). Transformed cell suspension cultures were selected, upscaled, harvested and protein extracts were used to perform IgG one-step Affinity Purification (AP) using metal-beads coupled to rabbit IgGs. (Figure 24 B).

Eluted proteins were then separated by SDS-PAGE and visualized by staining with Coomassie Blue (Figure 24 C). MRG1 with a GS tag fused at the C-terminal and N-terminal part and MRG2 with a GS tag fused at the C-terminal and the N-terminal part were analysed. The most putative interaction partners were identified when MRG1 and MRG2 were N-terminally fused to GS (Figure 24 C). Thus in the following results from these AP-MS experiments are discussed. Lists of interactors of MRG1-SG and MRG2-SG are shown in Supplementary Table 18. Proteins reproducibly identified in a free GS negative control were classified as unspecific interactors and were not included in the analyses. Additionally co-purified proteins were compared with a list of nonspecific interactors (Van Leene et al., 2015). GS-MRG2 and interacting proteins were also purified without the use of the enzyme benzonase, an endonuclease that digests DNA and RNA and that is used to exclude non-specific protein interactions that are mediated by RNA or DNA binding (Table 6) but any relevant differences compared to the results where benzonase was used came out. The Coomassie staining of the whole protein extracts (Input) resolved by SDS-PAGE was compared to the respective GS-bait eluate (Figure 24 C). The gel lanes with the AP eluates were cut into several pieces for in-gel tryptic digest and further subjected to mass spectrometry. Proteins which co-purified with the bait proteins were identified and were considered reliable only if present in at least two out of three AP-MS replicates. Only proteins with a MASCOT score higher than 80 and at least two peptides with an individual peptide score > 25 were further analysed (Figure 24 B).

### 3. Results: The chromatin status and alternative splicing



**Figure 24. MRG1/2 and PTB2 affinity purification coupled with mass spectrometry.** (A) Schematic illustration of PTB2 bait protein C-terminally fused to a SG tag and MRG1 and MRG2 N-terminally fused to a GS tag. Protein G: IgG-binding domain of protein G; TEV: tobacco etch virus (TEV) protease cleavage site; SBP, streptavidin-binding peptide. (B) Schematic workflow of AP purification of SG-tagged/GS-tagged proteins with IgG-coupled magnetic beads. (C) Total protein extracts of PSD-B cells expressing bait proteins. Proteins were separated by 9% SDS-PAGE and gels were stained with Coomassie Blue. Black asterisks indicate the bands corresponding to specific GS or SG-bait proteins. Numbers (A) indicate molecular weights in kDa computed by ExPaSy pl/Mw tool.

A variety of components of the splicing machinery like SR and hnRNP proteins co-purified with PTB2-SG what is in line with its role as a splicing regulatory factor shown in other studies (Barta et al., 2008; Reddy and Shad Ali, 2011; Ruhl et al., 2012b; Wachter et al., 2012). Unfortunately, MRG1 or MRG2 could not be detected as putative interactors of PTB2 (Table 2).

### 3. Results: The chromatin status and alternative splicing

**Table 2 Splicing-related proteins co-purifying with PTB2-SG.** The numbers indicate the respective average MASCOT scores and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

PTB2-SG	AGI	PROCESS	COMPLEX	INTERACTOR
23633/3	AT5G53180	Splicing	hnRNP family	PTB2
6264/3	AT1G20960	Splicing	U5 snRNP	EMB1507
1246/3	AT4G03430	Splicing	U5 snRNP	EMB2770
1068/3	AT2G43770	Splicing	U5 snRNP	WD40
537/3	AT2G33730	Splicing	U5 snRNP	RNA helicase 21
321/2	AT5G61140	Splicing	U5 snRNP	helicase
121/2	AT5G08290	Splicing	U5 snRNP	YLS8
1902/3	AT5G16780	Splicing	U4/U6.U5 tri snRNP	MDF
623/3	AT2G40650	Splicing	U4/U6.U5 tri snRNP	splicing factor 38A
458/2	AT3G05760	Splicing	U4/U6.U5 tri snRNP	U4/U6.U5 tri-snRNP
479/2	AT5G37370	Splicing	U4/U6.U5 tri snRNP	ATSRL1
634/3	AT1G60170	Splicing	U4/U6 snRNP	Prp31a emb1220
639/3	AT1G28060	Splicing	U4/U6 snRNP	atSAP90-1
586/3	AT2G41500	Splicing	U4/U6 snRNP	EMB2776
463/2	AT2G38730	Splicing	U4/U6 snRNP	atTri-20
737/2	AT1G60200	Splicing	U1snRNP related	RNA-binding protein 25
451/2	AT1G44910	Splicing	U1snRNP related	atPRP40A
186/2	AT3G07860	Splicing	U11/U12 specific	small nuclear ribonucleoprotein
1738/3	AT3G50670	Splicing	U1 snRNP	atU1-70K
896/3	AT5G17440	Splicing	U1 snRNP	atLuc7b
826/3	AT3G03340	Splicing	U1 snRNP	atLuc7a
354/3	AT5G51410	Splicing	U1 snRNP	atLuc7-ri LUC7 N
842/2	AT2G47580	Splicing	U1 snRNP	atU1A
385/2	AT4G03120	Splicing	U1 snRNP	atU1C C2H2
315/3	AT1G03910	Splicing	SR-related proteins	hypothetical protein
246/2	AT2G29210	Splicing	SR-related proteins	serine/arginine
934/3	AT3G49430	Splicing	SR proteins	AtSR34a
890/3	AT3G61860	Splicing	SR proteins	atRSp31
661/3	AT2G46610	Splicing	SR proteins	atRSp32
566/3	AT1G55310	Splicing	SR proteins	atSR33/atSCL33
574/3	AT3G55460	Splicing	SR proteins	atSCL30
556/3	AT4G31580	Splicing	SR proteins	atRSZp22/atSRZ22
459/3	AT1G23860	Splicing	SR proteins	atRSZp21/atSRZ21
339/3	AT5G18810	Splicing	SR proteins	atSCL28
430/3	AT3G13570	Splicing	SR proteins	AtSCL30a
451/3	AT4G35785	Splicing	SR proteins	RNA-binding protein

### 3. Results: The chromatin status and alternative splicing

<b>PTB2-SG</b>	<b>AGI</b>	<b>PROCESS</b>	<b>COMPLEX</b>	<b>INTERACTOR</b>
483/3	AT1G02840	Splicing	SR proteins	SR1
354/3	AT2G24590	Splicing	SR proteins	atRSZp22a
544/3	AT1G09140	Splicing	SR proteins	ATSR30
338/3	AT2G37340	Splicing	SR proteins	atRSZ33
164/3	AT5G64200	Splicing	SR proteins	atSC35
558/2	AT5G52040	Splicing	SR proteins	atRSp41
157/2	AT3G53500	Splicing	SR proteins	RSZ32
777/3	AT3G44850	Splicing	SR protein kinase	atSRPK2c
432/3	AT5G22840	Splicing	SR protein kinase	atSRPK2a
401/3	AT4G32660	Splicing	SR protein kinase	AME3
450/3	AT4G24740	Splicing	SR protein kinase	FC2
689/3	AT4G36690	Splicing	Splice site selection	atU2AF65a
763/3	AT1G60900	Splicing	Splice site selection	atU2AF65b
371/2	AT5G42820	Splicing	Splice site selection	atU2AF35
299/2	AT1G27650	Splicing	Splice site selection	atU2AF35a
840/2	AT1G02140	Splicing	EJC/mRNP	HAP1
187/2	At1g51510	Splicing	EJC/mRNP	Y14
3076/3	AT5G64270	Splicing	17S U2 snRNP	atSAP155
2066/3	AT3G55200	Splicing	17S U2 snRNP	atSAP130a
1568/3	At1g14650	Splicing	17S U2 snRNP	atSAP114-1a
1155/3	AT4G21660	Splicing	17S U2 snRNP	atSF3b150
664/3	AT5G06160	Splicing	17S U2 snRNP	atSAP61
492/3	AT2G18510	Splicing	17S U2 snRNP	atSAP49a
376/3	AT1G09760	Splicing	17S U2 snRNP	atU2A'
476/2	AT2G32600	Splicing	17S U2 snRNP	atSAP62
384/2	AT5G12190	Splicing	17S U2 snRNP	atP14-1
158/2	AT2G30260	Splicing	17S U2 snRNP	atU2B"b
1658/3	AT5G25060	Splicing	17S U2 associated	atSR140-1
542/3	AT1G30480	Splicing	17S U2 associated	atSPF45
665/2	AT2G47330	Splicing	17S U2 associated	atPrp5-2

Besides its role in splicing, PTB2 may also be directly connected to transcription as several transcription factors and transcription related proteins co-purified with PTB2-SG (Supplementary Table 15).

A putative interaction between PTB2 and MRG proteins was also not detected when MRG1 and MRG2 were used as bait proteins. PTB2 did not co-purify with MRG1 or MRG2 regardless of the orientation of the SG tag (Supplementary table 18).

In both MRG AP-MS experiments, MRG1 N-terminally or C-terminally and MRG2 N-terminally or C-terminally fused to a GS-tag, a variety of proteins related to splicing could be identified (Table 3, 4). Besides spliceosome components and SR proteins several core components of the MOS4-associated complex (MAC) (Prp19 complex in yeast, NTC complex in humans) could be detected.

**Table 3. Splicing-related proteins co-purifying with GS-MRG1.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

GS-MRG1	AGI	PROCESS	COMPLEX	INTERACTOR
2794/3	AT4G37280	Transcription	Histone acetyl transferase complex	MRG1
409/2	AT5G19300	Mtase	Mtase	C9orf114
337/3	AT2G43770	Splicing	U5 snRNP	atU5-40 Prp8 binding WD40 protein
1058/3	AT5G16780	Splicing	U4/U6.U5 tri snRNP	MDF
448/2	AT3G05760	Splicing	U4/U6.U5 tri snRNP	U4/U6.U5 tri-snRNP component
422/3	AT1G56290	Splicing	NTC-associated	CwfJ-like protein
795/2	AT2G38770	Splicing	NTC-associated	atAquarius EMB2765
365/3	AT3G58570	Splicing	EJC/mRNP	DEAD-box
502/2	AT1G77180	Splicing	Core NTC	SKIP
633/2	AT3G55200	Splicing	17S U2 snRNP	atSAP130a
328/3	AT2G18510	Splicing	17S U2 snRNP	atSAP49a
1570/2	AT1G20960	Splicing	U5 snRNP	atU5-200-2a EMB1507
377/2	At1g09770	Splicing	Core NTC	CDC5 (MOS4 PRL1 interactor)
420/3	AT4G21660	Splicing	17S U2 snRNP	atSF3b150
502/3	AT5G25060	Splicing	17S U2 associated	atSR140-1
987/3	AT5G64270	Splicing	17S U2 snRNP	atSAP155
117/2	AT5G06160	Splicing	17S U2 snRNP	atSAP61
727/2	AT3G62310	Splicing	RES complex	atPrp43-2a
614/3	AT1G18080	Splicing	Abundant first in C complex	WD-40
542/2	AT4G02840	Splicing	Sm core proteins	atSmD1-b
462/2	AT3G07590	Splicing	Sm core proteins	atSmD1-
356/2	AT1G18630	Splicing	Glycine-rich	GR-RBP6
324/2	AT4G08580	Splicing	B complex a	microfibrillar-

Besides a connection between MRG1/2 and splicing, analysis of the AP-MS results also revealed a possible interaction between MRG proteins and methyltransferase(s). In the GS-MRG2 AP-MS experiments the methyltransferases SDG7 (ASHH3) and C9orf114 were reproducibly co-purified (Table 4). C9orf114 was also detected as a putative interactor of GS-MRG1 (Table 3) and SDG7 co-purified reproducibly with MRG1-SG (Supplementary Table 18).

**Table 4. Splicing-related proteins co-purifying with GS-MRG2.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

GS-MRG2	AGI	PROCESS	COMPLEX	INTERACTOR
2198/3	AT1G02740	Chromatin reader	histone acetyl transferase complex	MRG2
525/3	AT2G44150	Mtase	Mtase complex	SDG7
1003/3	AT5G19300	Mtase	Mtase complex	C9orf114
2044/2	AT1G20960	splicing	U5 snRNP	atU5
571/2	AT2G33730	splicing	U5 snRNP	atU5-
555/2	AT4G03430	splicing	U5 snRNP	atU5
475/2	AT2G43770	splicing	U5 snRNP	atU5-40
135/2	AT5G09390	splicing	U5 snRNP	CD2 antigen
682/3	AT5G16780	splicing	U4/U6.U5 tri snRNP	MDF;
177/2	AT2G40650	splicing	U4/U6.U5 tri snRNP	splicing
399/3	AT1G28060	splicing	U4/U6 snRNP	atSAP90-1
391/2	AT2G41500	splicing	U4/U6 snRNP	atSAP60
133/2	AT5G51410	splicing	U1 snRNP	atLuc7-rl
387/2	AT3G49430	splicing	SR proteins	AtSR34a
98/2	AT5G64200	splicing	SR proteins	atSC35
754/2	AT1G09770	splicing	Core NTC	CDC5
640/2	AT5G28740	splicing	Core NTC	splicing
474/2	AT2G33340	splicing	Core NTC	MOS4-
410/2	AT1G04510	splicing	Core NTC	MOS4
393/1	AT3G18790	splicing	Core NTC	atlsy1
324/2	AT4G15900	splicing	Core NTC	PRL1
287/2	AT5G41770	splicing	Core NTC	atCRN1c
224/3	AT1G07360	splicing	NTC-associated	atECM2-1a
237/3	AT2G18510	splicing	17S U2 snRNP	atSAP49a s
876/2	AT3G55200	splicing	17S U2 snRNP	atSAP130a
451/2	AT4G21660	splicing	17S U2 snRNP	atSF3b150
362/2	AT1G14650	splicing	17S U2 snRNP	atSAP114-1a
711/2	AT5G25060	splicing	17S U2 associated	atSR140-1 U2

Proteins related to transcription were also identified when MRG1 and MRG2 were used as the bait protein (Supplementary table 16, 17). One subunit of RNAPII and some components of the PAF complex could be co-purified. IWS1 was identified only in one replicate of GS-MRG2 so it could not be classified as a true interactor.

Since no putative interactions between PTB2 and MRG proteins were detected by AP-MS experiments, it was additionally tested if an interaction of MRG proteins with splicing factors including PTB2 could be mediated by RNA or DNA. Thus GS-MRG2 APs were also performed without the use of the endonuclease benzonase. The obtained results (Table 5) did not show relevant differences compared to the results where benzonase was used (Table 4) and also no putative interaction between MRG2 and PTB2 could be detected.



**Table 5. Selected Proteins co-purifying with GS-MRG2 (AP without benzonase).** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

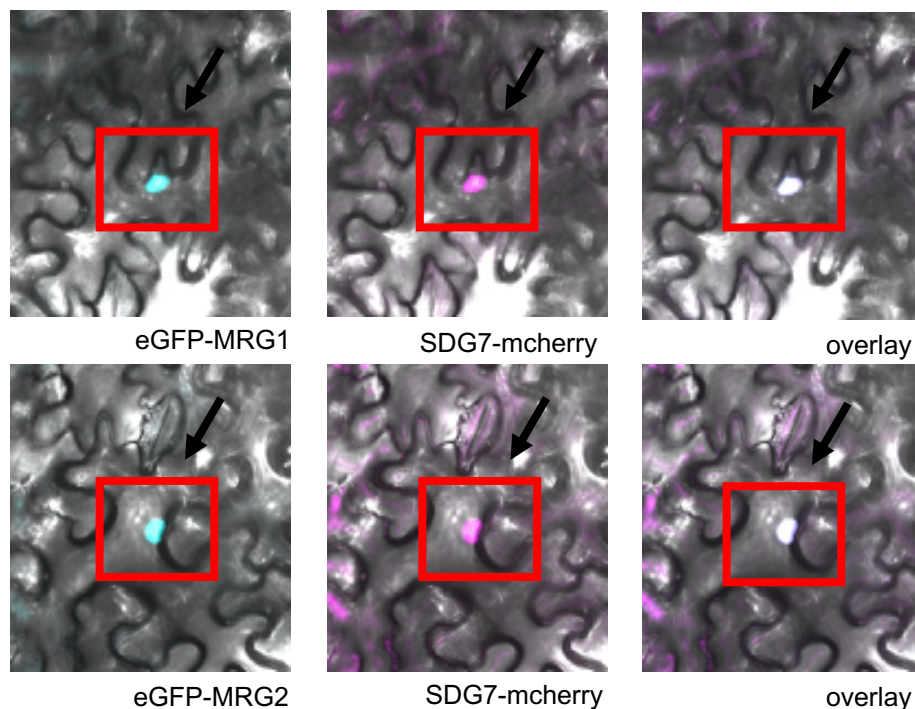
GS-MRG2	AGI	PROCESS	COMPLEX	INTERACTOR
3044/2	AT1G02740	Chromatin reader	histone acetyl transferase complex	MRG 2
845/2	AT5G19300	Metase	Metase complex	C9orf114
511/2	AT2G44150	Metase	Metase complex	SDG7
421/2	AT2G43770	splicing	U5 snRNP	atU5-40 Prp8
327/2	AT2G33730	splicing	U5 snRNP	atU5-100KD
1403/2	AT5G16780	splicing	U4/U6.U5 tri snRNP	MDF
441/2	AT3G05760	splicing	U4/U6.U5 tri snRNP	U4/U6.U5
134/2	AT1G04080	splicing	U1 snRNP	atPrp39a late flowering
146/2	AT3G61860	splicing	SR proteins	atRSp31
463/2	AT1G56290	splicing	NTC-associated	CwfJ-like protein
242/2	AT2G38770	splicing	NTC-associated	atAquarius EMB2765
181/2	AT1G07360	splicing	NTC-associated	atECM2-1a
920/2	AT2G27100	splicing	splicing	SERRATE
171/2	AT5G63120	splicing	mRNA binding	RNA helicase 30
389/2	AT3G26420	splicing	hnRNP family	ATRZ-1A
113/2	AT1G50300	splicing	hnRNP family	at-hnRNP-P
361/2	AT3G58570	splicing	EJC/mRNP	RNA helicase 52
547/2	AT1G77180	splicing	Core NTC	SKIP
140/2	AT2G33340	splicing	Core NTC	(MAC3B), PRP19B
193/2	AT3G55200	splicing	17S U2 snRNP	atSAP130a
136/2	AT2G18510	splicing	17S U2 snRNP	atSAP49a splicing factor
134/2	AT5G12190	splicing	17S U2 snRNP	atP14-1
471/2	AT5G04290	transcription	SPT4/SPT5	SPT5L
263/2	AT3G22320	transcription	Polymerase I, II, III, IV	NRP(A/B/C/D)5
170/2	AT4G29830	transcription	PAF-C	SKI8, VIP3
883/2	AT1G26470	chromatin remodelling	SWR1/NuA4	EAF7
618/2	AT2G47210	chromatin remodelling	SWR1/NuA4	AtSWC4
314/2	AT3G24870	chromatin remodelling	SWR1/NuA4	AtEAF1B

### 3.3 MRG proteins directly interact with the methyltransferase SDG7

To examine whether MRG proteins directly interact (protein-protein interaction, PPI) with the methyltransferases C9orf114 and/or SDG7 (ASHH3) the Matchmaker TM GAL4 yeast-two hybrid (Y2H) system and FRET experiments (Förster resonance energy transfer) were applied to test for a direct protein-protein interaction. After no PPI between MRG proteins and C9orf114 could be detected by Y2H (experiment performed by BSc Tobias Schnekenburger 2018, with the use of

positive and negative control, data not shown) a possible PPI between SDG7 and MRG proteins was analysed by Y2H and Förster resonance energy transfer (FRET).

Before experiments analysing putative PPIs between SDG7 and MRG proteins were conducted, the subcellular localizations of an SDG7-mCherry and MRG1/2-eGFP fusion proteins were determined. The coding sequences of *MRG1* or *MRG2* and *SDG7* fused to the respective fluorophore were co-expressed under the 35S CaMV promoter in *N. benthamiana* leaves after *A. tumefaciens* mediated transformation. CLSM showed that all three proteins are uniformly distributed in the nucleus (Figure 25).



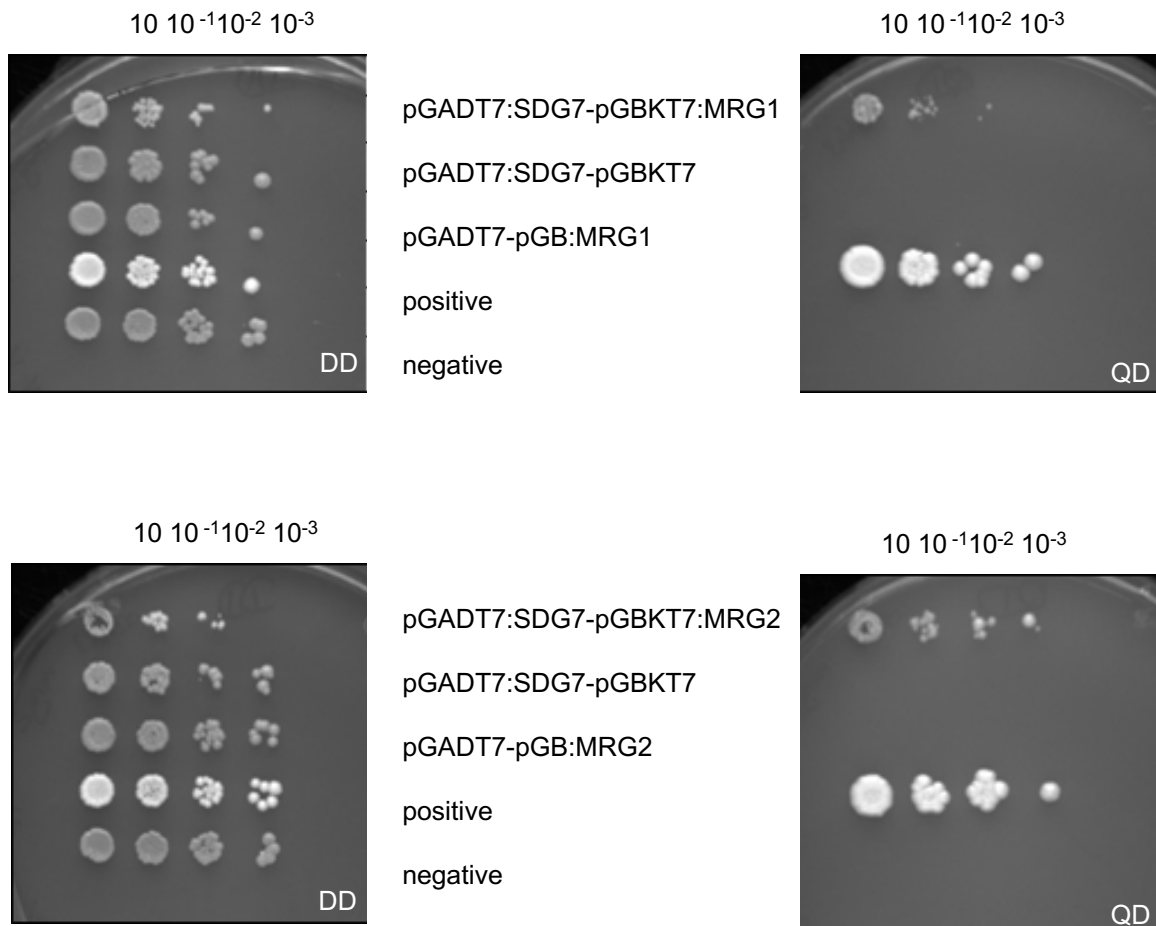
**Figure 25. MRG proteins and SDG7 are localized to the nucleus.** Co-expression of eGFP-MRG1/MRG2 (cyan) and SDG7-mCherry (magenta) in the epidermal cells of *N. benthamiana* leaves.

In a first approach to test for PPI between MRGs and SDG7, the Y2H system was used. MRG proteins fused to the DNA-binding domain (DNA-BD, bait) of GAL4 and SDG7 fused to the activation domain (AD, prey) of GAL4 were expressed in the auxotrophic yeast strain AH109. Yeast cells co-expressing both types of fusion proteins were scored for their growth. Whereas all transgenic strains expressing both bait and prey plasmids grew on the double dropout (DD) medium lacking leucine (-Leu) and tryptophan (-Trp), only the strains containing interacting proteins grow on the high-stringent selective quadruple (QD; -Leu/-Trp/-His/-Ade) dropout plates (Figure 26). Cells expressing DNA-BD/murine p53 and AD/SV40 large T-antigen served as a positive control exhibiting normal growth on DDO and QD plates (Iwabuchi et al., 1993). Cells expressing DNA-BD/Lamin and AD/SV40 large T-antigen were used as a negative control with no growth on the QD media. As an additional negative control, cells co-expressing DNA-BD/MRG2 or MRG1 with AD/ SV40 large T-antigen and DNA-BD/murine p53 with AD/SDG7 were studied showing no background detection on QDO plates.

### 3. Results: The chromatin status and alternative splicing

Yeast cells expressing both the AD-SDG7 and MRG1 or MRG2-BD fusion proteins grew on the high-stringent, selective quadruple dropout (QD) medium, demonstrating that SDG7 interacts with MRG1 and MRG2 in yeast cells (BSc Anja Pavlica, 2018). Strains expressing only one of the fusion proteins and the negative control showed no growth on the QD medium (Figure 26).

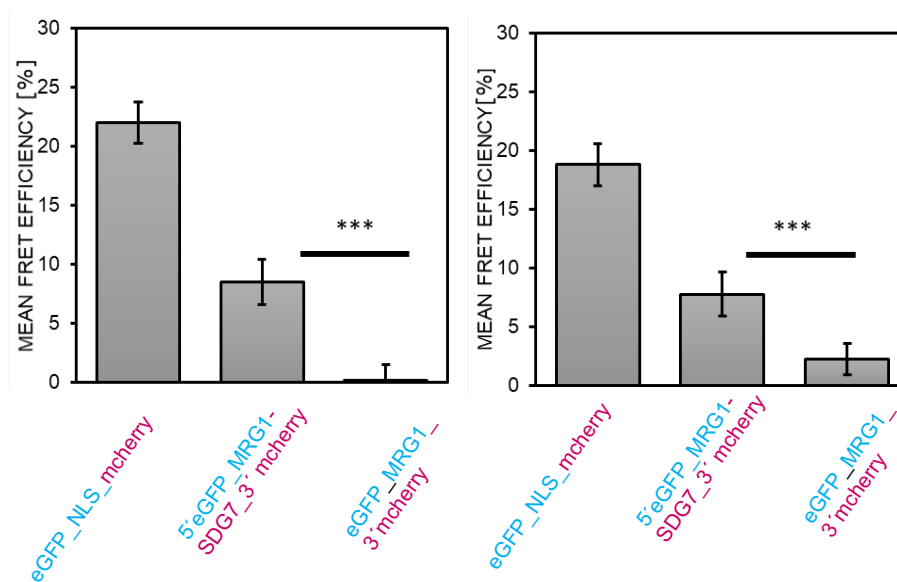
Also, PTB2 and IWS1 proteins were tested by Y2H for a possible PPI with MRG1 and MRG2 but no interaction could be detected (data not shown).



**Figure 26. MRG1 and MRG2 protein directly interact with SDG7.** Direct protein interaction was tested by Y2H. MRGs fused to the DNA-binding domain (BD) of GAL4 were expressed in yeast cells, whereas the SDG7 sequence was fused to the activation domain (AD) of GAL4. All strains grew on the double dropout (DD) medium (left panel) whereas only the strains expressing both, the AD-SDG7 and MRG1, MRG2-BD fusion proteins grew on the high-stringent, selective quadruple dropout (QD) medium.

The interaction between SDG7 and MRG1/2 was further analysed by Förster resonance energy transfer (FRET) in *A. tumefaciens* infiltrated *N. benthamiana* leaves. Co-transformed epidermal cells were identified by confocal microscopy and subjected to FRET analysis. The efficiency of the energy transfer was measured using the Acceptor Photo-Bleaching method (Weidtkamp-Peters and Stahl, 2017). The coding sequences of *MRG1*, *MRG2* or *SDG7* were fused to *eGFP* (*MRG1* and *MRG2*) and *mCherry* (*SDG7*) and donor/acceptor pairs were expressed by the strong CaMV 35S promoter.

The recovery of donor fluorescence after acceptor photobleaching (APB) was quantified. FRET measured in cells expressing an eGFP-mCherry fusion protein served as a positive control for the experiment while FRET measured in cells co-expressing MRG1/2-eGFP with unfused mCherry served as a negative control. In the positive control eGFP is physically linked to mCherry that ensures a transfer of energy from one fluorescent protein to the other that can be measured as a mean FRET-APB efficiency. Mean FRET-APB efficiencies of approximately 22% and 19% were measured for the positive control in the experiments where an interaction of MRG1 or MRG2 with SDG7 was tested (Figure 27) what is similar to other FRET experiments performed in our lab (Ehrnsberger et al., 2019; Pfaff et al., 2018). For both tested interactions (MRG1-eGFP-SDG7-mCherry and MRG2-eGFP-SDG7-mCherry) a significant higher mean FRET-APB efficiency could be measured when MRG1-eGFP or MRG2-eGFP was co-expressed with SDG7-mCherry (~ 7.5% mean FRET-APB efficiency) compared to the co-expression of MRG1-eGFP or MRG2-eGFP and free mCherry (Figure 27) what confirms a direct protein interaction between both MRG proteins and SDG7.



**Figure 27. Protein interactions analysed by FRET experiments.** *N. benthamiana* leaves were coinfiltrated with vectors (x axis) directing the expression of the indicated donor (eGFP, blue)-acceptor (mCherry, red) combinations as well as of positive and negative controls. Transiently transformed cells were analysed in three biological replicates by FRET-APB. Mean FRET-APB efficiencies ( $\pm$ sd, eight analysed nuclei each) are shown, which were analysed using Tukey's test (p-value < 0.05) using Past3 (Hammer et al., 2001).

### 3.4 Molecular characterization of *mrg* and *iws1* T-DNA insertion lines

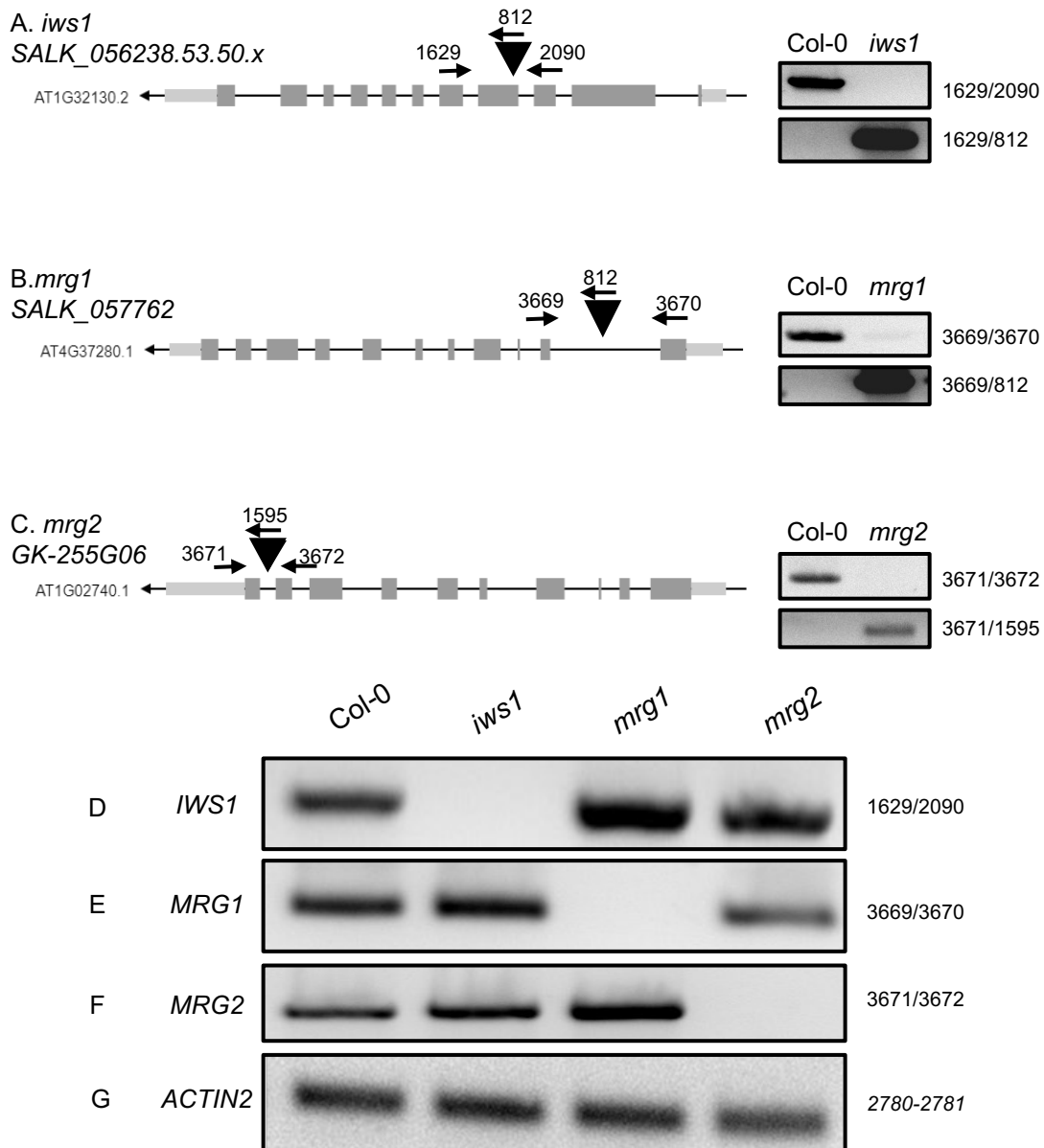
In a next step to characterize the IWS1 – H3K36me3 – MRG adaptor complex in *Arabidopsis*, the influence of IWS1 and MRG proteins on plant growth and development was examined. Therefore a

double mutant plant lacking MRG1 and MRG2 was generated and phenotypically analysed. In mammals it could be demonstrated that IWS1 recruits the methyltransferase SDG8 which deposits the histone mark onto chromatin. H3K36me3 on the other hand is the target for the chromatin reader protein MRG15 which subsequently recruits splicing factors (Sanidas et al., 2014). Even though IWS1 was not identified as a putative direct interactor of MRG proteins in the AP experiments (Table 5,6, supplementary table 16, 17, 18), a possible connection between IWS1 and MRG proteins was further analysed. Therefore a homozygous triple mutant plant lacking IWS1, MRG1 and MRG2 was generated and phenotypically analysed. The T-DNA line IWS1 was already present in our lab and it is the same line used in RNAseq and ChIPseq experiments.

T-DNA lines were obtained from the Nottingham Arabidopsis Stock Centre (NASC). A schematic overview of the three T-DNA lines used in this study is depicted in Figure 31. Genotyping PCR was done with one primer pair spanning the T-DNA insertion and one primer pair with one primer binding to the T-DNA and one primer binding to a neighbouring DNA region (Figure 28). The T-DNA insertion positions are indicated in the schematic representations of the genes by triangles. All three T-DNA lines were then made homozygous for the T-DNA insertions tested by PCR based genotyping (Figure 28 A-C).

To determine if the T-DNA insertions in the three genomic sequences reduce the transcript levels of the respective full-length mRNAs, semi-quantitative RT-PCR was performed (Figure 28 D-G). RNAs were extracted from 14 DAS seedlings and converted to cDNA. Actin 2 was used as reference gene (Figure 28 G). Comparing the relative levels of the *IWS1* transcripts between the different genotypes shows that *IWS1* is altered in *iws1*, but not in *mrg1* and *mrg2* mutants (Figure 28 D). *MRG1* and *MRG2* transcripts on the other hand cannot be detected in *mrg1* or *mrg2* respectively but no downregulation can be detected in the other genotypes (Figure 28).

Overall, this shows that T-DNA insertions in the homozygous mutant lines clearly reduce the transcript levels of the respective full-length mRNAs.



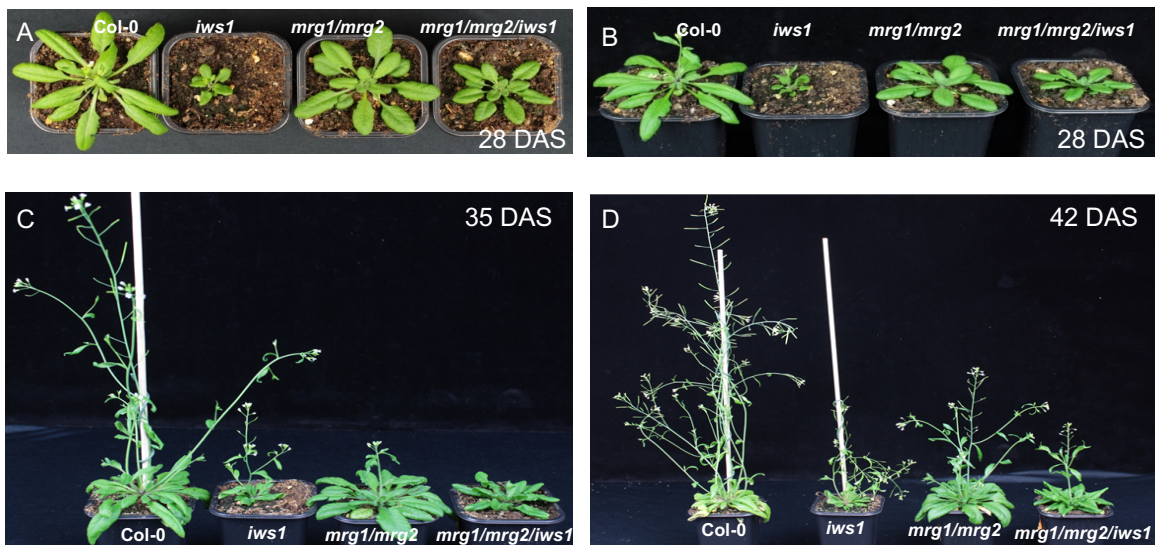
**Figure 28. Molecular characterization of *iws1-2*, *mrg1* and *mrg2* T-DNA lines.** Schematic representations of *IWS1* (A) *MRG1* (B) and *MRG2* (C) with site of the T-DNA insertion shown by triangles and primers used for PCR based genotyping and semi quantitative RT-PCR. (D) Expression analysis. Semi quantitative RT-PCR on *IWS1* (D), *MRG1* (E), *MRG2* (F) and *ACTIN2* (G) as a reference gene. RT-PCR was performed on cDNA made from RNA of Col-0 and the three single mutants.

### 3.5 Phenotypic analysis of *mrg* and *iws1* mutant lines

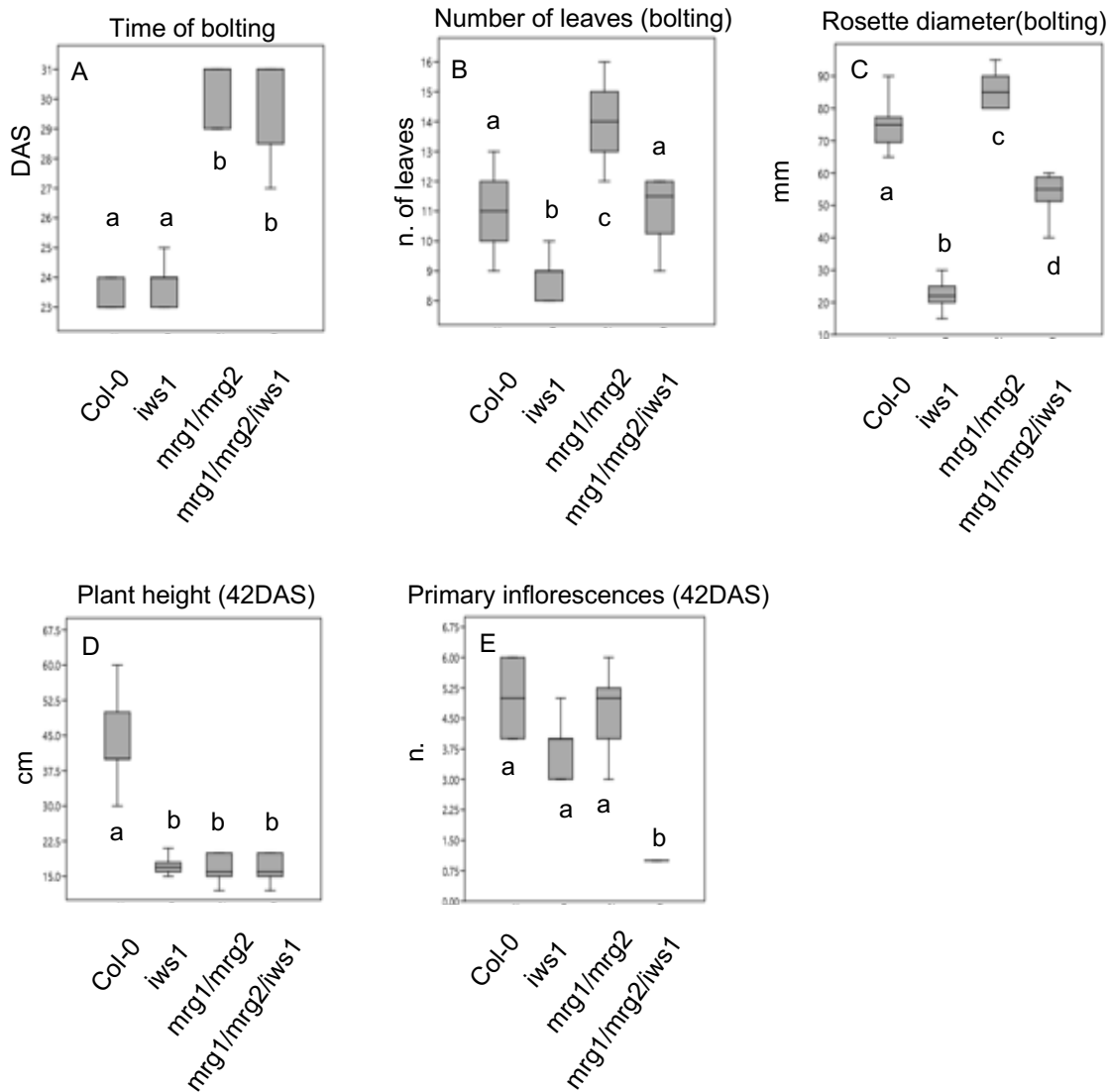
For the phenotypic analysis plants were grown on soil under long-day conditions and developmental and morphological stages were monitored at the following time points: time of bolting, 28 and 42 days after stratification (DAS).

When looking at the time of bolting, *mrg1/mrg1/iws1* triple mutant plants show the same late flowering phenotype like *mrg1/mrg2* double mutant plants, whereas the depletion of *IWS1* alone does not result in a change in flowering time compared to Col-0, suggesting an epistatic genetic interaction (Figure 29 A,D, Supplementary Table 13). The number of leaves and rosette diameter at bolting on

the other hand indicates that there is an additive genetic interaction between MRG proteins and IWS1 as the triple mutant plants show a statistically significant intermediate phenotype between *iws1* and *mrg1/mrg2* plants (Figure 30 A,B, Supplementary Table 13). Later in plant development at 42 DAS, *mrg1/mrg2/iws1* triple mutant plants show a similar plant height like the two other mutants what does support a synergistic genetic interaction between the factors (Figure 29 C, 30 C, Supplementary Table 13, Supplementary Figure 7). Primary inflorescences at this stage are heavily reduced in the triple mutant plants compared to *iws1*, *mrg1/mrg2* and Col-0 suggesting ones more a synergistic genetic interaction between MRGs and IWS1 in regulating the building of primary inflorescences (Figure 29 C, 30 E).



**Figure 29.** Phenotypic analysis of *iws1 mrg1/mrg2* and *mrg1/mrg2/iws1* compared to Col-0. Representative individuals are shown at various developmental stages. A,B) 28 DAS, C) 35 DAS and D) 42 DAS.



**Figure 30. Phenotypic analysis of *iws1 mrg1/mrg2* and *mrg1/mrg2/iws1* compared to Col-0.** Phenotypic evaluation with (A) time of bolting (Elongation of the first internode), (B) Number of leaves at bolting, (C) Rosette diameter at bolting, (D) Plant height 42 DAS and (E) primary inflorescences 42 DAS. All data are means (n=15) ± SD, the box plots indicate the outcome of a multi comparisons Tukey's test (p-value < 0.05). Box plot and statistical analysis were made with *Past3* (Hammer et al., 2001).

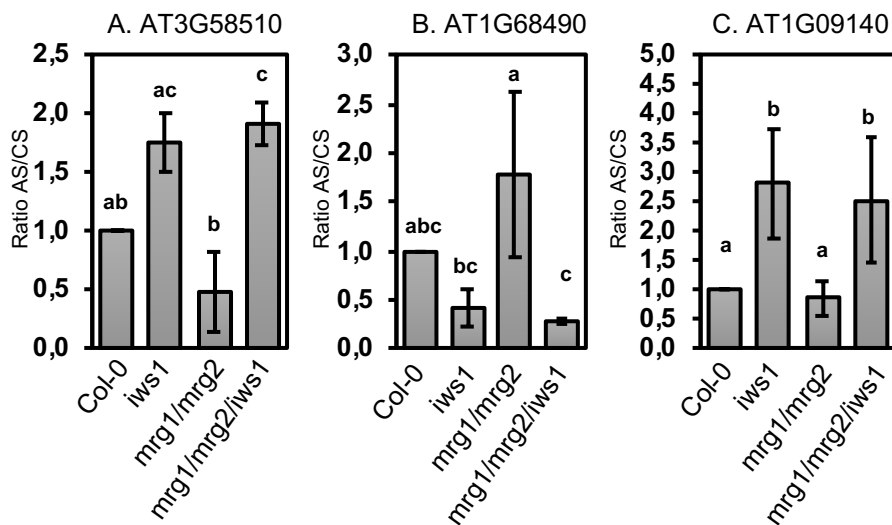
### 3.6 MRG proteins do not influence AS pattern of selected AS genes

Since several splicing factors were found to interact with MRG proteins (Table 4, 5), it was further tested if MRGs are involved in the regulation of alternative splicing. To analyse if the loss of MRG proteins affects AS, splicing pattern of three alternatively spliced genes were analysed in the *iws1*, *mrg1/2* and *mrg1/mrg2/iws1* mutant lines compared to Col-0. For the experimental validation two candidate genes (At3g58510 and At1g68490) that were identified by RNAseq to be differentially alternatively spliced in *iws1* (Table1, Figure 14) were selected. Additionally one SR gene (SR30,



At1g09140) which is known to be alternatively spliced and which is often used to test for alternative splicing (Palusa et al., 2007) was included.

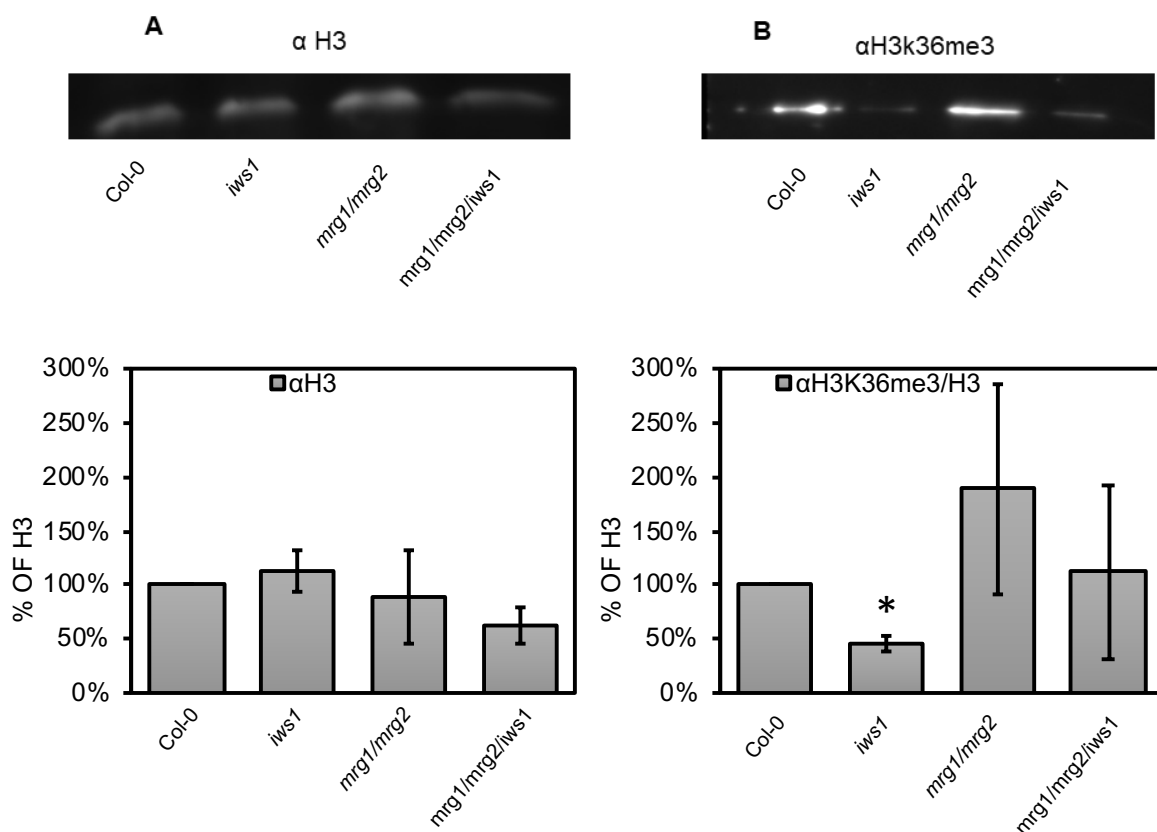
RNAs were isolated from the different mutant lines and Col-0 and cDNAs were generated. Primers were used to amplify by RT-PCR fragments from the alternatively spliced (AS) isoform and the constitutively spliced isoform (CS) and the quantity of amplified DNA fragments was assayed using a Bioanalyzer (Supplementary Table 19, Supplementary Figure 8). No changes relative to Col-0 were observed in the ratios of splice variants in *mrg1/mrg2* in the three genes analysed (Figure 31). On the other hand for all three tested genes a significant different AS/CS ratio was detected in *mrg1/mrg2/iws1* compared to *mrg1/mrg2* (Figure 31). Regarding the splicing ratio quantified in the single *iws1* mutant the amount of alternatively spliced isoform compared to Col-0 was consistent with the RNAseq validation experiment in chapter 2.4, Figure 14.



**Figure 31. AS of three selected genes in *iws1* and *mrg* mutant plants.** Major and minor spliced isoforms of 3 candidate genes were amplified by RT-PCR and quantity of amplified fragments was determined by bioanalyzer measurements. Mean ratios of alternatively spliced isoform (AS)/constitutively spliced isoform (CS) from three independent biological replicates normalized to Col-0 are shown ( $\pm$  SD). Different letters between genotypes show statistically significant differences in the ratios of AS event/ CS event between these genotypes determined by Tukey's test (P-value < 0.05) using Past3 (Hammer et al., 2001)

### 3.7 Global levels of H3K36me3 are reduced in *iws1* mutant plants

In the following it was analysed if also in *Arabidopsis* the TEF IWS1 is involved in the regulation of H3K36me3 like observed in human cells (Sanidas et al., 2014). In a first approach the global levels of H3K36me3 were determined by Western Blot in the mutant lines *iws1*, *mrg1/mrg2* and *mrg1/mrg2/iws1*. For the quantification antibodies against H3K36me3 were used and the measured levels were normalized to the global H3 levels. Compared to Col-0, the global H3K36me3 levels were significantly reduced in *iws1* and also in the triple mutant *mrg1/mrg2/iws1* a clear but not significant reduction in H3K36me3 levels compared to *mrg1/mrg2* could be detected.



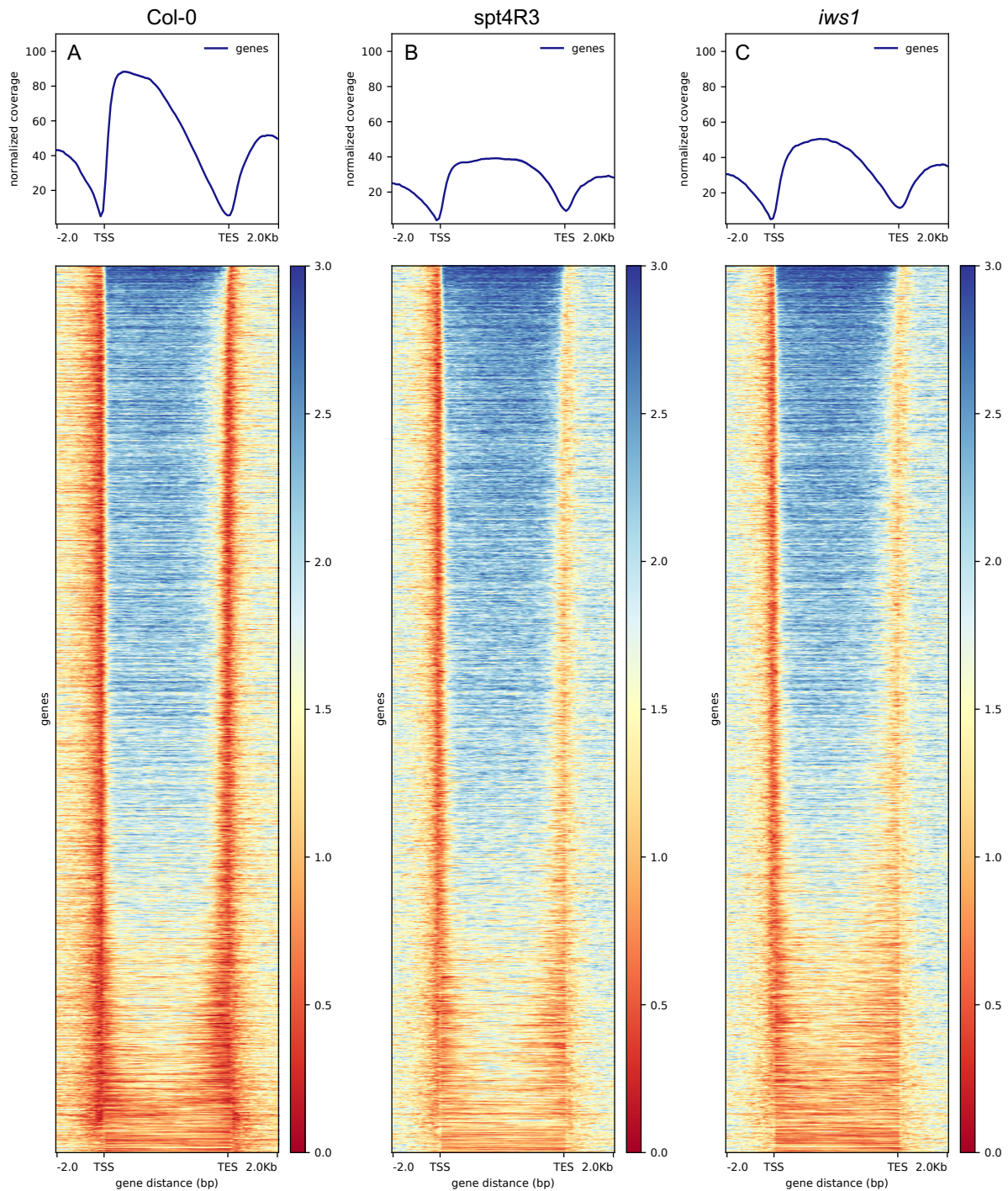
**Figure 32. Western blot analysis of global H3K36me3 levels.** (A, B) Immunoblot with αH3 (A, control) and (B) αH3K36me3 antibodies using whole protein extracts from 14 DAS Col-0, *iws1*, *mrg1/mrg2*, and *mrg1/mrg2/iws1* seedlings. Band intensities detected by Western Blot (B) were measured by ImageJ and normalized to band intensities of (A). Error bars indicate SD of three independent experiments. The star above the histogram bars indicate the outcome of a multi comparisons Tukey's test (p-value < 0.05) using Past3 (Hammer et al., 2001).

### 3.8 Global H3K36me3 density is reduced in *tef* mutant plants

To analyse the global impact of TEFs on the deposition of H3K36me3 and to address the question whether this histone mark might be involved in the regulation of AS in *Arabidopsis*, ChIPseq analysis in *iws1* and *spt4R3* mutant plants using H3K36me3 antibodies was performed. ChIPseq was performed on two biological replicates that showed a high reproducibility. H3K36me3 distribution was determined over all TAIR10 annotated genes and was depicted from 2 kbp upstream the transcription start site (TSS) to 2 kbp downstream the transcription end site (TES). The bioinformatic analyses were performed in our group by Simon Obermeyer.

Global analysis showed that in Col-0 the H3K36me3 distribution shows the highest density immediately downstream of the TSS before the H3K36me3 distribution gradually decreases towards the TES (Figure 33). In both mutant lines H3K36me3 levels were clearly reduced downstream the TSS. Additionally in both mutant lines towards the TES the drop in H3K36me3 levels was not as pronounced.

### 3. Results: The chromatin status and alternative splicing

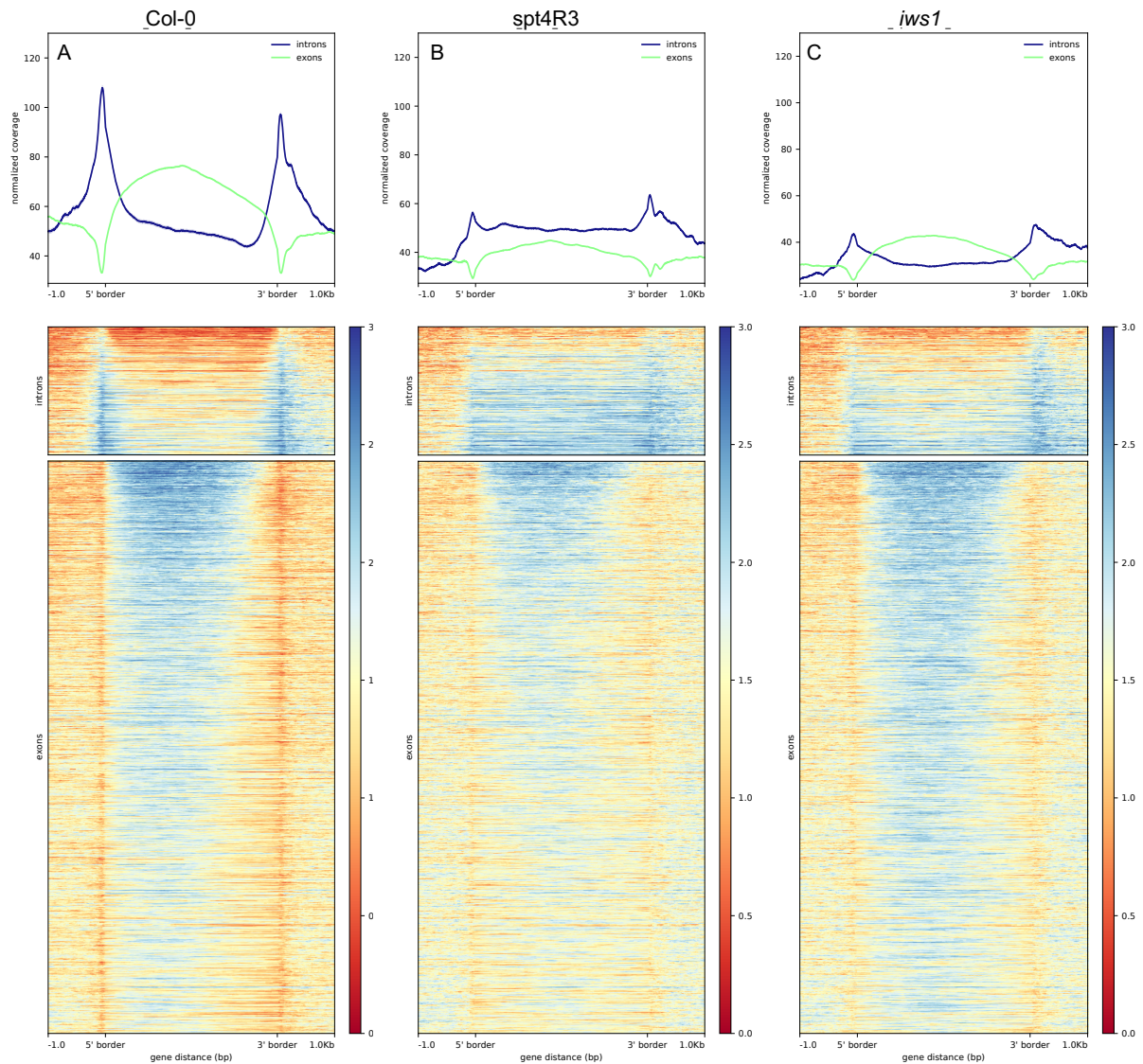


**Figure 33.** H3K36me3 density over all TAIR 10 annotated genes determined by ChIPseq. H3K36me3 densities in Col-0 (A), *spt4R3* (B) and *iws1* (C). H3K36me3 densities are displayed in the range between 2 kb upstream of the TSS and 2 kb downstream of the TES. All ChIP-seq tracks are normalized against the according Input libraries. Figures were created by Simon Obermeyer.

To analyse the H3K36me3 distribution along exons, introns and at intron/exon boundaries, all introns and exons were isolated from the TAIR10 genome annotation and additionally filtered for introns and exons that are larger than 500bp. In the following all introns and exons were scaled to the same length (5' border to 3' border) and the H3K36me3 distributions over introns and exons were analysed (Figure 34). In Col-0 the H3K36me3 distribution was higher along exons than along introns and a

### 3. Results: The chromatin status and alternative splicing

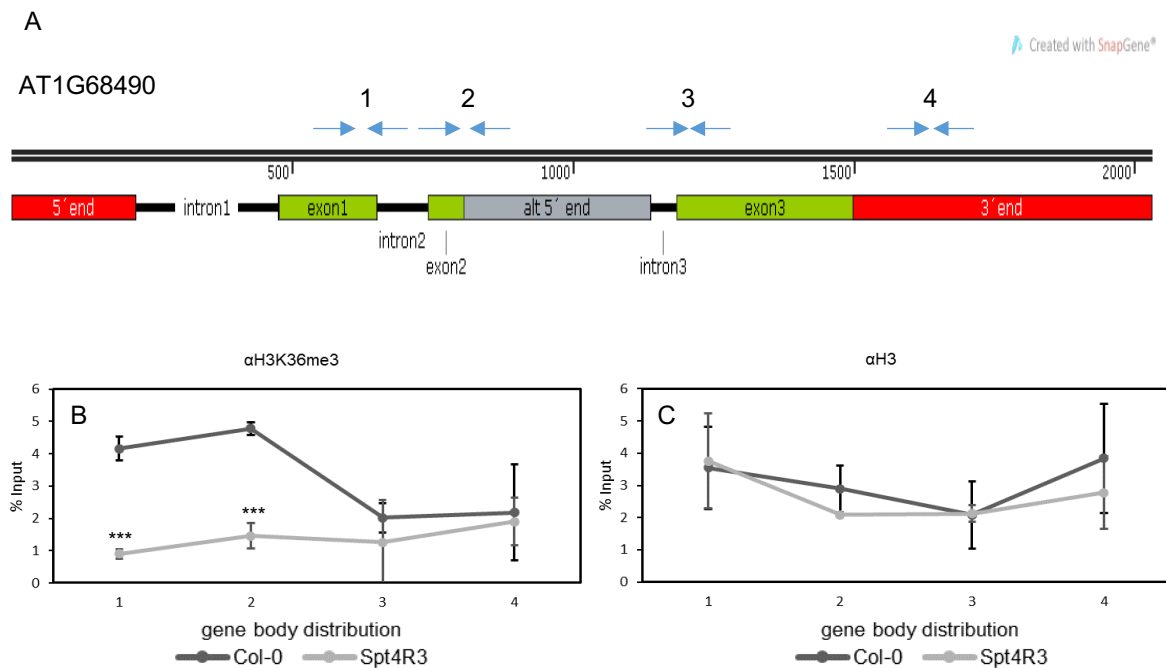
clear and distinct enrichment around intron/exon boundaries could be observed (Figure 34 A). In both mutant lines H3K36me3 exon marks were clearly reduced. Additionally the enrichment of H3K36me3 at intron/exon boundaries was clearly not as pronounced in *iws1* and *spt4R3*.



**Figure 34.** Mean H3K36me3 distribution over exons and introns measured by ChIP-seq. Density line plots and heat map reads of A) Col-0, B) *spt4R3* and C) *iws1*. Introns and exons were isolated from the TAIR10 genome annotation and filtered for introns and exons that are larger than 500bp. Introns (blu lines) and exons (green lines) were scaled to the same length (5' border to 3' border). Figures were created by Simon Obermeyer.

### 3.9 Validation of H3K36me3 ChIPseq results by ChIP-qPCR

By ChIPseq it was demonstrated that globally H3K36me3 levels are enriched around the TSS and that these histone marks decrease gradually towards the 3' end. In both mutant lines at the TSS the levels were reduced compared to Col-0 and the drop in H3K36me3 marks was less pronounced at the TES (Figure 33, Supplementary Figure 10). In the following the H3K36me3 distribution was determined in *spt4R3* and *iws1* mutant plants at an individual locus by ChIP-qPCR along At1g68490, a gene that is differentially alternatively spliced in both mutant lines (Table1, Figure 14) and which was also used to experimentally validate the S2P and S5P ChIPseq results (Figure 21, 22).

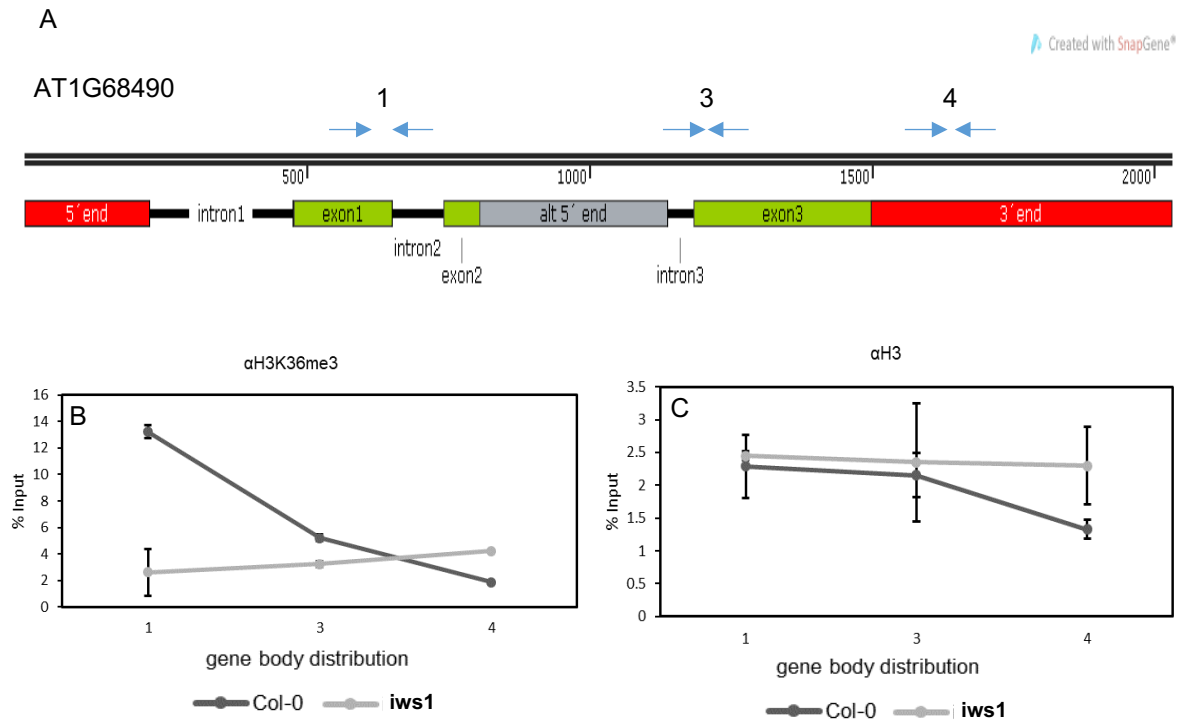


**Figure 35 ChIP-qPCR validation experiment comparing the H3K36me3 distribution over At1g68490 in Col-0 and *spt4R3*.** A) Schematic representation of At1g68490 with the AS event r(5' SS) depicted in grey. Green boxes represent exons and red box 5'UTR and 3'UTR. The arrows above indicate the relative position of primer pairs (1-4) used for qPCR. B,C) qPCR analysis after chip using H3 (B) and H3K36me3 (C) specific antibodies. Results are displayed as percentages of the input as determined by qPCR (Y-axis). Error bars represent  $\pm$  SD of at least two biological and two technical replicates. X-axis displays primer pairs (A) used for qPCR..

In mutant line *spt4R3* qPCR validation using primer pair 1 (Figure 35 A), which amplifies a region in exon1, showed that H3K36me3 levels are significantly reduced compared to Col-0 in this region (Figure 35 B). A significant reduction can also be observed more downstream around the constitutive 5' SS (primer pair 2) before H3K36me3 levels converge in Col-0 and *spt4R3* towards the 3' end (primer pairs 3, 4). A reduction of H3K36me3 marks in *spt4R3* near the 5' end of At1g68490 is also indicated by the ChIPseq track of this gene (Supplementary Figure 9).

Since in this experiment differences in the distribution were mainly observed around the regions amplified with the primer pair combinations 1, 3 and 4, it was decided to continue with these combinations in the follow validation analysis. In the mutant line *iws1* a similar H3K36me3 profile with a significant reduction at exon1 (Figure 36 A, primer pair 1) of At1g68490 can be observed

(Figure 36 B). Towards the 3' end also in *iws1* the H3K36me3 levels converge compared to Col-0 (primer pair 3,4).



**Figure 36** ChIP-qPCR validation experiment comparing the H3K36me3 distribution over At1g68490 in Col-0 and *iws1*. A) Schematic representation of At1g68490 with the AS event (5' ss) depicted in grey. Green boxes represent exons and red box 5'UTR and 3'UTR. The arrows above indicate the relative position of primer pairs (1,3,4) used for qPCR. B,C) qPCR analysis after chip using H3 (B) and H3K36me3 (C) specific antibodies. Results are displayed as percentages of the input as determined by qPCR (Y-axis). Error bars represent  $\pm$  SD of at least two biological and two technical replicates. X-axis displays primer pairs (A) used for qPCR.

The results obtained could confirm the genome-wide distribution observed by ChIP-seq and the RNAseq validation analysis.

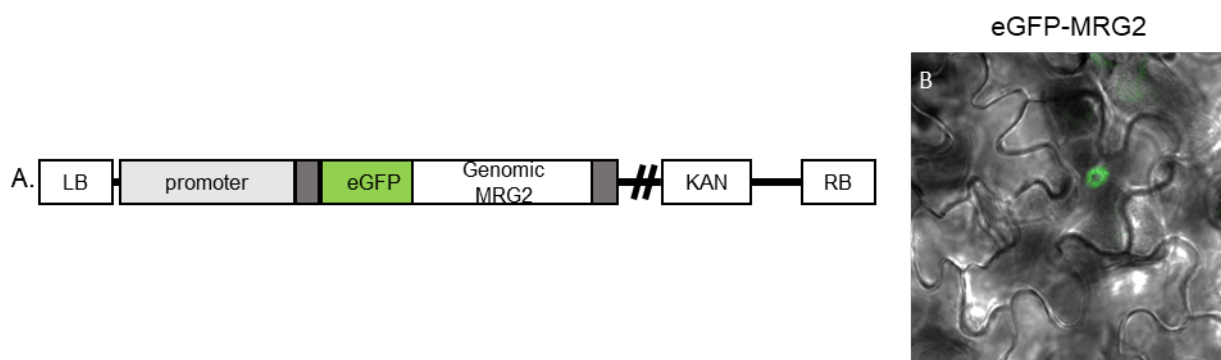
Analysing the single gene ChIPseq profile, a reduction of H3K36me3 marks in *iws1* near the 5' end of At1g68490 is also indicated by the ChIPseq track of this gene (Supplementary Figure 9).

### 3.10 Recruitment of MRG2 in *iws1* mutant plants

RNAseq and ChIPseq in *iws1* plants indicated that IWS1 is involved in the regulation of AS and that this TEF also contributes to the regulation of the H3K36me3 histone deposition. In order to describe the putative plant IWS1 – H3K36me3 – MRG adaptor complex and to analyse if this system regulates AS like the human counterpart (Sanidas et al., 2014) it will be necessary to globally analyse whether MRG binding to H3K36me3 is dependent on IWS1. This can be achieved by ChIP of MRG proteins in Col-0 and *iws1* mutant plants and a comparison of chromatin binding of MRGs and H3K36me3 distribution.

Since MRG2 co-purified with more splicing related interaction partners than MRG1 (Table 4,5), it was decided to perform a ChIP experiment with MRG2 to test whether MRG2 binding to chromatin is dependent on IWS1. For that purpose *Arabidopsis* plant lines stably expressing *eGFP-MRG2* under its native promoter were generated. The introduced transgene consists of a 6.57 kb genomic *MRG2* region N-terminally fused to *eGFP* (Figure 37 A). Primers for the amplification of the native promoter were taken from a previous publication (Xu et al., 2014). Before the vector was introduced into Col-0 and *iws1* mutant plants, *eGFP-MRG2* protein expression and subcellular localization was first analysed in *A. tumefaciens* infiltrated *N. benthamiana* leaves. Confocal laser scanning microscopy (CLSM) examination of the transformed cells expressing pMRG2::eGFP-MRG2 allowed detection of green fluorescence in nuclei (Figure 37 B). The primers used to amplify the native promoter in the follow experiments were designed according to the publication from Xu et al., 2014. Our findings show that MRG1 and MRG2 are localised at euchromatin regions in the nucleus what is consistent with a previous report (Xu et al., 2014).

Transgenic *Arabidopsis* lines were generated by *A. tumefaciens*-mediated transformation. Three independent lines homozygous for the *eGFP-MRG2* construct were identified by resistance to the antibiotic kanamycin as well as by genotyping PCR.

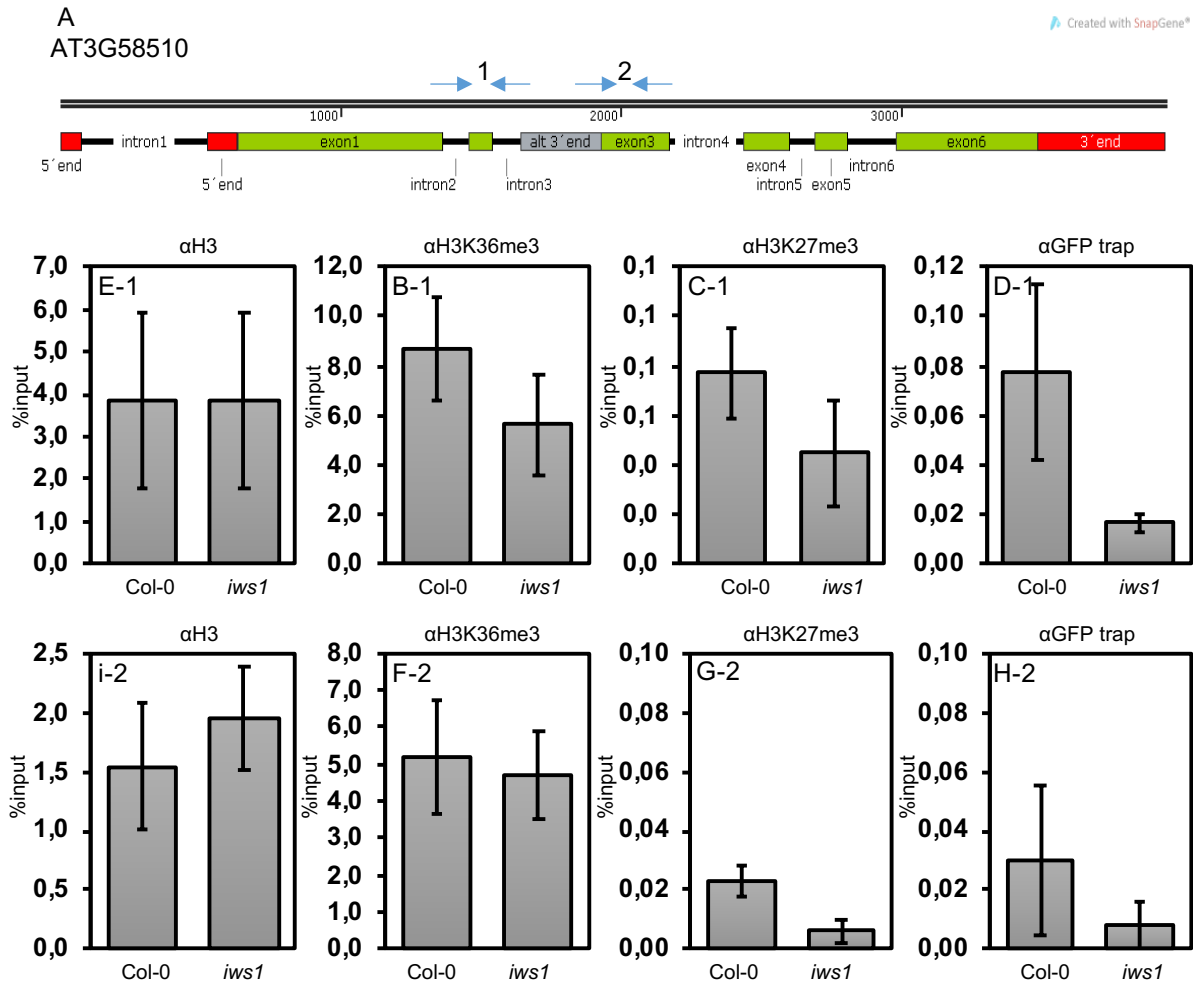


**Figure 37. MRG2 is localized in the nucleus.** (A) Schematic illustration of the *eGFP-MRG2* transgene construct. The sequence of *eGFP* is inserted in the genomic locus of *MRG2* as N-terminal translational fusion (KAN resistance marker, light grey bar = promoter, genomic *MRG2* sequence with intron and exon, green bar = *eGFP* coding sequence, dark grey bars = UTRs, LB/RB = right/left border). (B) The representative *eGFP-MRG2* line was subjected to confocal microscopy to investigate the subcellular localization of the *eGFP*-tagged protein in *Arabidopsis* epidermal leaf.



To test if the recruitment of MRG2 to the chromatin might be controlled by IWS1, a ChIP in Col-0 and *iws1* mutant plants using antibodies against eGFP-MRG2 (GFP-TRAP®) was performed. Furthermore to shed light on the role of histone modifications in the recruitment of MRG proteins to chromatin, ChIP experiments using antibodies against H3K36me3 and H3K27me3 were conducted. QPCR with primer pairs binding near the AS event in the differentially alternatively spliced gene AT3G5810 (Table 1) was used to determine the density of histone modifications and MRG2 binding at the respective region (Figure 38 A).

The histone H3 densities in the analysed genomic regions were similar in Col-0 and *iws1* supporting the integrity of the results (Figure 38 B,F). When using antibodies against H3K36me3 and H3K27me3 a higher density of these histone marks could be detected in the more upstream region (primer pair 1), a pattern that was also detected in the ChIP experiment using *iws1* mutant plant material (Figure 37). Both histone marks displayed a reduced density on both analysed regions in the *iws1* mutant compared to Col-0 whereas the reduction was more pronounced for the H3K27me3 histone mark (Figure 38 C,G). Using GFP-TRAP® antibodies for ChIP also a reduced binding of eGFP-MRG2 to both analysed regions of AT3G5810 could be observed in *iws1* mutant plants compared to Col-0 (Figure 38 H,D). Thus a correlation between a reduced H3K36me3 and H3K27me3 density and a reduced MRG2 binding to the respective regions could be shown for the AT3G5810 gene. The experiment clearly showed a tendency that MRG2 recruitment is regulated by IWS1 but for definitive conclusions more replicates and other genes need to be tested in upcoming experiments.



**Figure 38. qPCR of At3g58510 after Chip experiment.** (A) Schematic representation of At3g58510 with the AS event represented in grey. Green boxes are exons and red boxes respectively 5'UTR and 3'UTR. The arrows above show the primers used for qPCR (1-2). (B) CHIP analysis of At3g58510 using antibodies against H3K36me3, H3K27me3 and the GFP-Trap in Col-0 and *iws1* mutant both expressing eGFP-MRG2. For CHIP experiments percentage of Input was determined by qPCR. Error bars indicate  $\pm$ SD of at least two biological and two technical replicates.

## 4. Discussion: The influence of transcription elongation factors on alternative splicing

Alternative splicing is a regulatory mechanism that modulates gene expression in various ways and that adds another layer of complexity to modulate transcriptome diversity. In mammalian cells, almost all protein-coding genes can undergo AS and organisms regulate a variety of essential functions by this mechanism (Gallego-Paez et al., 2017). Although the extent is less, also plants use AS to fine-tune their physiology and metabolism in normal and in stress conditions but the molecular mechanism how AS is regulated is poorly understood (Jabre et al., 2019).

Accumulating evidence in humans indicates the functional coupling between transcription elongation and mRNA maturation processes like mRNA splicing (Saldi et al., 2016). In plants it was recently demonstrated that splicing also occurs mainly co-transcriptionally. In 2016 depleted intron reads could be observed by GRO-seq (Global nuclear run-on sequencing) what indicated the existence of co-transcriptional splicing also in *Arabidopsis* (Hetzl et al., 2016). Antosz et al., 2017 co-precipitated a variety of splicing factors and some components of the spliceosome with the RNAPII TEC, suggesting that the mRNA processing machinery associates at different stages with the RNAPII TEC. In 2020 the co-transcriptional nature of splicing could also be demonstrated globally and it was shown that nearly all introns undergo co-transcriptional splicing in *Arabidopsis* (Li et al., 2020).

The coupling of transcription and splicing implies that the elongating RNAPII may also be involved in the regulation of the splicing process. Two models are proposed to explain how the elongating RNAPII influences the splicing decision and consequently AS (Brody et al., 2011; Dvinge, 2018; Jimeno-González et al., 2015; Luco et al., 2011). The recruitment model states that the transcription elongation complex (TEC) directly or indirectly recruits splicing factors by an interaction with the CTD of RNAPII (De La Mata and Kornblihtt, 2006) Furthermore the nucleosome density and/or the chromatin status is involved in the recruitment of splicing factors to RNAPII (Kornblihtt et al., 2004; Luco et al., 2010; Saldi et al., 2016; Zhou et al., 2014). The kinetic model on the other hand proposes that the elongation rate of RNAPII influences the splicing decision in a way that a slow RNAPII allows additional time for upstream weak splice sites to recruit splicing factors before downstream stronger splice sites emerge during pre-mRNA synthesis (Dujardin et al., 2014; Fong et al., 2014). In *Arabidopsis* multiple evidences support the kinetic model. Interestingly, pharmacologic and genetic experiments showed that RNAPII elongation speed is faster under light conditions than in darkness what correlates with a change in AS of a subset of genes (Godoy Herz et al., 2019a; Petrillo et al., 2014). A connection between open chromatin, a higher RNAPII speed and AS further demonstrated that the RNAPII elongation rate is linked to the choice of alternative splice sites by the splicing machinery (Ullah et al., 2018).

In yeast, experiments indicated that besides the chromatin environment also the assembly of certain transcription elongation factors (TEFs) with RNAPII may be important for an efficient transcription elongation (Mayer et al., 2010). In *Arabidopsis*, elongating RNAPII is also associated with various conserved TEFs, indicating that also in plants the assembly to a functional TEC may be important for RNAPII elongation speed and consequently alternative splicing decisions (Antosz et al., 2017).

The development of genome wide sequencing and omics approaches over the last decade made it possible to also globally shed light on alternative splicing in plants (Filichkin et al., 2010; Marquez et al., 2012; Zhang et al., 2017). In this study genome wide analyses were performed to examine the impact of TEFs on co-transcriptional splicing since still relatively little is known about this connection in *Arabidopsis*.

## 4.1 Different TEFs affect alternative splicing of different targets

The impact of TEFs absence on AS was examined by RNAseq in the *tef* mutant plant lines *spt4R3*, *iws1*, *tflls*, as well as in the double mutant line *tflls/iws1*. Data analysis showed that the loss of these factors results in a significant change in AS for a variety of genes.

RNAseq analysis on AS genes under different light conditions triggered rapid AS responses of numerous genes in etiolated *Arabidopsis* seedlings (Hartmann et al., 2016). Also a change in ambient temperature conditions results in a significant amount of differentially alternative spliced genes in *Arabidopsis* (Pajoro et al., 2017). The number of genes differentially alternatively spliced in the *tef* mutants was smaller compared to the AS analyses using different temperatures or stress conditions indicating that the analysed TEFs induce splicing of only a subset of genes.

Strikingly, the overlap of differentially alternatively spliced genes was not pronounced between the different mutant lines with only 6.5% percent of the detected differentially alternatively spliced genes being shared by all three analysed single mutant genotypes. This indicates that in *Arabidopsis* the different TEFs influence AS of distinct subset of genes or the observed effects are caused indirectly. A similar pattern could be observed when analysing the impact of TEFs on global gene expression. The loss of specific TEFs results in the misregulation of only subsets of genes (Van Lijsebettens and Grasser, 2014). The downregulation of *SPT4* for instance resulted in the differential expression of only ~5% of the genes and genes involved in auxin signalling were overrepresented in this group (Dürr et al., 2014; Van Lijsebettens et al., 2014).

The least genes (32) were alternatively spliced in *tflls*. In yeast it could be demonstrated that the TFIIIS orthologue is dispensable for cells grown under normal conditions and the expression of TFIIIS enhances exon skipping (Fish and Kane, 2002; Svejstrup, 2007, Howe et al., 2003). Also in *Arabidopsis* the inactivation of the single-copy *TFIIS* gene does not affect plant growth and only ~2.3% of the genes are differentially expressed in a *tflls* mutant line (Van Lijsebettens and Grasser, 2014). These findings indicate a rather minor impact of TFIIIS on cell viability and development in yeast and *Arabidopsis* what correlates with the comparatively small impact of TFIIIS on AS. The key function of TFIIIS is to stimulate RNAPII to reassume elongation after pausing (Fish and Kane, 2002). The non-essential character of TFIIIS indicates that also the unstimulated RNAPII can transcribe genes (Fish and Kane, 2002; Svejstrup, 2007). The introduction of a dominant negative TFIIIS version on the other hand results in various developmental and morphological defects in *Arabidopsis* (Antosz et al., 2020; Dolata et al., 2015). Besides these phenotypic alterations also AS of a variety of genes is affected in this mutant line and most of the observed changes correspond to what would be predicted by the kinetic model (Dolata et al., 2015; Shukla et al., 2011). Together with the data from this study these findings may indicate that the minor impact of TFIIIS on AS correlate with a possible minor impact of TFIIIS on the transcription elongation rate and that TFIIIS is more likely not directly involved in the recruitment of splicing related factors.

The impact of *SPT4* and *IWS1* on AS was to some extent higher. In the *spt4R3* RNAi mutant line 57 genes and in the *iws1* T-DNA line 93 genes were identified as differently alternatively spliced. These stronger alterations in AS can be correlated with a severely reduced growth of *spt4R3* and *iws1* plants and one can speculate that also misregulated AS contributes to these mutant plant phenotypes.

*SPT4* as part of the heterodimeric *SPT4/5* complex supports the elongation stage of RNAPII transcription by modulating pausing or arrest during transcript elongation (Hirtreiter et al., 2010; Van Lijsebettens and Grasser, 2014; Wada et al., 1998). Several studies in yeast and metazoan systems

#### 4. Discussion: The influence of transcription elongation factors on alternative splicing

also showed that SPT4/5 is involved in splicing but if this influence is due to its capacity to directly interact with the RNAPII to couple chromatin modification states and RNA processing or if it directly recruits splicing factors to the TEC needs further evaluation (Hartzog and Fu, 2013). The data from this work add evidences to previous studies (Burckin et al., 2005; Lindstrom et al., 2003), where it already has been postulated that SPT4 may play a role in coordinating splicing with transcription by influencing RNAPII kinetic elongation rate and/or the spliceosome assembly.

The loss of IWS1 has the biggest impact on AS among analysed TEFs. This can be explained with the fact that IWS1 might regulate splicing in various ways. In yeast and mammalian systems it could be demonstrated in vitro through a targeted genetic screen that IWS1 (Spn1 in yeast) is involved in chromatin remodelling and histone modifications and by that it could influence RNAPII elongation rate through the methylation of H3K36, which is a mark of transcriptionally active chromatin (Yoh et al., 2008; Zhang et al., 2008). Additionally, histone modifications regulated by IWS1 can also recruit splicing factors to chromatin what adds another layer of splicing regulation related to IWS1 (Luco et al., 2010). In line with that, several studies in mammalian cells showed that IWS1 is involved in mRNA splicing (Oqani et al., 2019; Yoh et al., 2007, 2010). A recent publication further indicated that also in *Arabidopsis* besides a possible regulatory role in RNAPII kinetics, IWS1 is also directly involved in the spliceosomal cycle (Kanno et al., 2020).

The data obtained in this study were based on experiments using T-DNA lines and it is possible that results were affected by indirect effects. To give a final conclusion about the biological relevance of TEFs on AS additional experiments must be performed.

## 4.2 The absence of TEFs affects alternative splicing of splicing regulators

Within the alternative spliced genes in the *tef* mutant plant lines, several genes encoding SR proteins were identified (Table 6).

**Table 6.** SR proteins identified in the different *tefs* to be differentially alternatively spliced

ATG	Description	Spt4R3	tflls	iws1	tflls/iws1	AS event
AT3G61860	RS31	3,5E-04	5,1E-02	8,5E-06	1,9E-05	intron_retention
AT4G31580	SRZ-22	3,9E-03	3,0E-02	3,3E-02	2,1E-02	intron_retention
AT3G49430	SR34a, SRp34a, At-SR34a	2,3E-01	4,1E-01	3,9E-04	5,2E-02	intron_retention
AT5G25280	SR protein	9,6E-01	8,7E-01	6,1E-04	3,7E-02	alt_5prime
AT1G02840	SR34	1,7E-01	3,7E-01	2,9E-02	7,0E-03	alt_5prime

It is worth to mention that among the differentially alternatively spliced genes detected by RNAseq some of the 'seven indicator genes' used in the first screening approach were found. These RNA binding proteins are extensively linked with alternative splicing (Reddy and Shad Ali, 2011). They can directly or indirectly interact with C-terminal domain of RNAPII (De La Mata and Kornblihtt, 2006; Sapra et al., 2009) The SR proteins are splicing enhancers and recruiters of splicing activator proteins, increasing the probability that a nearby site will be used as a splice junction (Reddy and

Shad Ali, 2011). It is also well documented that genes encoding SR proteins are alternatively spliced and that organisms regulate splicing decisions by that pathway (Jeong, 2017; Zhou and Fu, 2013). Additionally, some SR proteins interact with H3 tails and associate with chromatin (Loomis et al., 2009). Besides SR proteins also other RNA binding proteins like the Gly-rich RNA binding protein At-GRP7 and the PTB homologues can act in auto- and cross- regulatory negative feedback loops by triggering splicing of their own pre-mRNA or that of related proteins in *Arabidopsis* (Staiger et al., 2003; Stauffer et al., 2010). Furthermore the SR protein SRSF2 mediates the release of paused RNAPII activating p-TEFb (Ji et al., 2013). The fact that genes encoding SR splicing factors are among the discovered alternatively spliced genes in the *tef* mutants further supports the theory that TEFs have an impact on the splicing outcome. In particular by their function in recruiting the splicing machine to the transcription complex and by regulating the proper assembly of the transcriptional and co-transcriptional machineries (Jonkers and Lis, 2015).

### **4.3 Intron retention is the most prevalent AS event altered in the *tef* mutants**

Notably, intron retention (IR) was the most prevalent AS event identified in the different *tef* mutant lines. This is in line with previous studies in plants where intron retention was detected as the most prevalent AS event (Drechsel et al., 2013b; Filichkin et al., 2015; Godoy Herz et al., 2019b; Mancini et al., 2016; Marshall et al., 2016; Pajoro et al., 2017; Seo et al., 2011), whereas exon-skipping is the prevalent AS event detected in human cells (Wang et al., 2008).

These differences might be explained with plant specific characteristics like a small number of introns per gene, long exons versus short introns and the presence of a large number of SR proteins (Reddy, 2007a). Furthermore, in mammals, the mechanism of splicing occurs predominantly in an “exon definition way”, initially recognizing an adjacent splice site of an exon (De Conti et al., 2013). In contrast, in plants, splice site recognition occurs predominantly by an “intron definition way”, (Reddy, 2007b) and intron features like intron number and intron position were shown to correlate with co-transcriptional splicing efficiency in *Arabidopsis* (Li et al., 2020; Zhu et al., 2020).

IR frequently generates transcripts harbouring a premature termination codon (PTC), which are degraded by the cytoplasmic nonsense-mediated decay (NMD) quality control pathway (Chaudhary et al., 2019b). But NMD not only functions in the clearance of erroneous transcripts, it also enables gene regulation via degradation of AS variants (Lareau et al., 2007). In *Arabidopsis*, AS coupled to NMD plays a major role in regulating the transcriptome and a study indicated that ~43% of all AS events produce NMD targets (Wang and Brendel, 2006). The biological significance of IR was further demonstrated since IR is involved in the regulation of several process like plant growth and development, flowering time and stress response (Drechsel et al., 2013; Filichkin et al., 2015, Marshall et al., 2016; Seo et al., 2011). The prevalence of this AS event in the analysed *tef* mutant lines may contribute to the observed plant phenotypes, like dwarfism of *iws1* plants, but this hypothesis needs further evaluation. IR transcripts are often sequestered in the nucleus under a particular stress or developmental stage for further processing upon cell requirement or degraded by the NMD pathway (Chaudhary et al., 2019b). Interestingly, in plants and mammals, NMD responses are reduced under stress conditions and it was proposed that plants may buffer against normal protein synthesis levels by AS to produce the protein isoforms needed for adaptation to stresses (Chaudhary et al., 2019b; Shaul, 2015). For example, inhibition of NMD mediates plants defence against pathogen attack by producing more salicylic acid (Rayson et al., 2012).

## 4.4 Different TEFs affect alternative splicing of different targets with similar biological functions

Although the overlap of genes alternatively spliced in the different *tef* mutants is rather small, striking similarities regarding enriched GO terms can be observed between genotypes. This indicates that the different TEFs are involved in the regulation of different genes but with similar biological function. In all three analysed single mutant lines the majority of enriched GO terms belong to the class metabolic processes. Analyses of the functional categories of genes associated with light-regulated AS in *Arabidopsis* revealed an overrepresentation of the term “metabolism” for blue light, what further indicates that genes which regulate metabolic processes are subjected to AS (Hartmann et al., 2016). Genes differentially spliced in *sdg8-2* and *sdg26-1* mutants under different ambient-temperature conditions are overrepresented by GO categories related in the metabolic process (Pajoro et al., 2017).

The representation of the “RNA” category in the GO term analysis is in line with previous publications (Filichkin et al., 2010; Rühl et al., 2012; Drechsel et al., 2013), showing that genes involved in RNA metabolism are extensively regulated by AS. In *Arabidopsis* after light stress AS patterns of genes were affected that belong to GO categories associated with RNA processing (Mancini et al., 2016). Furthermore light-regulated AS revealed an overrepresentation of the term “RNA” for blue light and white light (Hartmann et al., 2016). Gene ontology (GO) term analysis for the loci associated with the Polypyrimidine Tract Binding protein (PTB1/2) regulated AS events also revealed overrepresentation of the term “RNA processing”. This suggests the existence of PTB-dependent regulatory networks among genes involved in RNA metabolism (Rühl et al., 2012).

The GO term stress response is also overrepresented in all mutant genotypes. It is well established that plants regulate stress response by AS (Chaudhary et al., 2019b; Jabre et al., 2019). For instance studies analysing the impact of different nutrient conditions on AS showed that differentially spliced genes were involved in regulatory mechanisms depending on the type of stress (Nishida et al., 2017).

Only small differences among mutant genotypes could be detected regarding overrepresented GO terms. The biological process of splicing is only overrepresented by candidate genes differentially alternatively spliced in *spt4R3* and *iws1* but not in *tfl1s*, indicating that SPT4 and IWS1 might be involved in the regulation of splicing pattern of splicing-related genes. Although this might at first appear as an unexpected outcome, splicing of splicing-related genes can potentially influence the composition of the spliceosome and subsequently alter the splicing of downstream targets, like genes that function in the circadian clock or light response of the plant (Filichkin et al., 2015; Marshall et al., 2016; Verhage et al., 2017).

## 4.5 The absence of TEFs influences RNAPII occupancy along genes

The development of new generation sequencing methods like NET-Seq and GRO-seq (Nojima et al., 2015; Zhu et al., 2018) have disclosed new important aspects in the relation between transcription elongation and AS. According to the kinetic and recruit model, RNAPII elongation speed and the recruitment of factors influence the splicing outcome and it can be speculated that TEFs may be involved in the regulation of RNAPII activity. A key role in the regulation and coordination of transcription initiation, elongation and other RNA processing activities is played by the phosphorylation status of the RNAPII CTD (Egloff and Murphy, 2008; Harlen and Churchman, 2017).

#### 4. Discussion: The influence of transcription elongation factors on alternative splicing

Especially Ser2 and Ser5 phosphorylations were shown to be important for plant transcription and are involved in stress responses and the regulation of plant development (Hajheidari et al., 2013). Besides its role in regulating the processivity of RNAPII, the phosphorylation status of the CTD also facilitates the recruitment of proteins involved in splicing (Alexander et al., 2010; Fusby et al., 2015; Hajheidari et al., 2013). In mammalian cells for instance the spliceosome is coupled with a Ser5 phosphorylated CTD but not with the Ser2 phosphorylated one (Nojima et al., 2015, 2018). Accordingly, in *Arabidopsis* an accumulation of Ser5P RNAPII at 5' splice sites accompanied by the recognition of the 5' splice site and the recruitment of the spliceosome could be observed (Zhu et al., 2018).

To analyse the global influence of TEFs absence on RNAPII activity and to examine a possible connection with AS, ChIPseq was performed in the *tef* mutant lines *iws1* and *spt4R3* and RNAPII profiles were generated. The results obtained by ChIPseq correlated with the transcriptome analysis indicating a cross-regulation between splicing and the chromatin landscape.

The results discussed in the following are based on only one biological replicate. Thus it is possible that the results do not display the biological coherences precisely and it will be necessary to confirm the obtained results in upcoming experiments..

In both mutant lines as well as in Col-0, Ser2P spreads over gene bodies and accumulates at the transcription end site (TES) what is consistent with the findings of a previous study (Antosz et al., 2020; Hetzel et al., 2016; Zhu et al., 2018). ChIPseq further revealed in both analysed mutant lines a higher Ser2P RNAPII occupancy along the gene bodies. This could indicate that the TEFs IWS1 and SPT4 might be involved in the regulation of the phosphorylation status of the CTD what could at the end affect the recruitment of the splicing machinery leading to altered AS. Ser2P RNAPII density was strongly enriched in the *spt4R3* mutant line whereas only a minor enrichment could be detected in the *iws1* mutant. This indicates that the TEF SPT4 may have a bigger impact on RNAPII activity than IWS1 and/or SPT4 may regulate Ser2P. Additionally, a more pronounced accumulation at the TES in the *spt4R3* mutant line further indicates that in *Arabidopsis* the TEF SPT4 may be involved in the transcription termination. The analysis on individual loci by ChIP-qPCR showed an enrichment consistent with earlier analyses (Ding et al., 2011; Dürr et al., 2014). The distribution of Ser2P along transcribed regions was similar to the Ser2P distribution observed in yeast and mammals (Hajheidari et al., 2013; Harlen and Churchman, 2017; Jeronimo et al., 2016).

The Ser5P RNAPII profiles show some distinct differences along gene bodies compared to the Ser2P RNAPII profiles. Whereas Ser2P RNAPII occupancies spread over gene bodies and show the highest accumulation around the TES, Ser5P RNAPII densities are enriched immediately downstream the TSS and the occupancies along the gene bodies get reduced before the TES. This pattern could also be observed in mammalian and yeast cells (Hajheidari et al., 2013; Harlen and Churchman, 2017; Jeronimo et al., 2016). In *Arabidopsis* this accumulation around the TSS was also detected in a previous study (Nojima et al., 2015, 2018; Zhu et al., 2018). Notably, in the *iws1* mutant line the ChIPseq profile indicates a higher Ser5P density along the gene bodies, a feature that was also described in yeast (Zhang et al., 2008). Additionally a defined peak at the TES can be observed what points to an IWS1-dependent regulatory mechanism of Ser5P in *Arabidopsis*. In contrast, the Ser5P RNAPII profile in the *spt4R3* mutant line is not different compared to Col-0.

The RNAPII profiles indicate that the loss of the TEFs IWS1 and SPT4 results in an increased density of RNAPII what might be caused by a slower RNAPII leading to altered AS in the *iws1* and *spt4R3* mutant lines. Additionally, the discrepancy that the loss of SPT4 results in an enrichment of Ser2P RNAPII and the loss of IWS1 results in an enrichment of Ser5P RNAPII along gene bodies may



explain why different targets are differentially alternatively spliced in the *spt4R3* and *iws1* mutant lines. According to the recruitment model the phosphorylation status of the CTD is involved in the recruitment of splicing factors. Thus Ser5P RNAPII and Ser2P RNAPII may recruit different factors leading to the splicing of different targets. Thus our data does not point to one model but it rather supports the idea that both mechanisms (kinetic coupling and recruitment coupling) may be involved in the regulation of AS.

## 4.6 RNAPII phosphorylated at Ser5 accumulates at intron/exon boundaries

In eukaryotes, RNAPII is not evenly distributed among exons and introns and a dynamic phosphorylation status of the CTD was shown to be coupled to the recognition of splice sites at intron/exon boundaries (Saldi et al., 2016). In human cells a higher RNAPII occupancy and a slower elongation rate were determined in exons than in introns (Saldi et al., 2016). Additionally, mNETseq revealed that splice sites are potent RNAPII pausing signals with Ser5 hyperphosphorylated RNAPII accumulating at 3' and 5' splice sites (Mayer et al., 2015; Nojima et al., 2015). A higher Ser2P and Ser5P RNAPII occupancy along exons than along introns was also detected by ChIPseq in this study, suggesting that also in plants a higher RNAPII density along exons accompanied by a slower elongation rate may be important for the regulation of splicing. Furthermore increased occupancy at 3' and 5' splice sites was also identified, displayed by an enrichment of Ser2P and especially Ser5P RNAPII around 3' and 5' borders of introns, indicating a similar pausing mechanism like in human cells. These results are in contrast to findings of a previous publication which showed that in *Arabidopsis* pausing mainly occurs at the 5'SS (Zhu et al., 2018). These discrepancies may be explained by the different methods that were used in the two studies to analyse RNAPII occupancies.

## 5. Discussion: The H3K36me3 – MRG adaptor complex

After a connection between TEFs, RNAPII elongation and AS could be demonstrated, it was focused on another regulatory mechanism of AS. According to the kinetic model, also nucleosome packaging and the chromatin landscape influence the RNAPII elongation rate. Some histone modifications, like for example acetylation, favour a more open chromatin state resulting in a faster RNAPII (Zhou et al., 2014), leaving the spliceosome less time to recognise weaker splice sites. On the other hand when the nucleosome density is high, the transcription elongation rate is slower (Chen et al., 2010; Fong et al., 2014; Tilgner et al., 2009) and also weaker splice sites are recognized leading to AS.

Besides their role in the kinetic model, histone modifications are also involved in the regulation of AS by recruiting splicing factors (recruitment model). In this study the attention was focused on the H3k36me3 modification and H3k36me3 readers capable to recruit splicing factors. This model is well studied in mammals where it could be demonstrated that by this regulatory mechanism the splicing of the human fibroblast growth receptor 2 (FGFR2) gene is controlled (Luco et al., 2010).

The H3K36me histone mark is read by the MORF-related gene 15 (MRG15) chromatin reader protein that can recruit the splicing factor PTB to the exon IIIb of FGFR2 (Sanidas et al., 2014; Luco et al., 2010).

## 5.1 MRG proteins co-purify with splicing factors

In *Arabidopsis* MRG1 and MRG2 bind H3K36me3 *in vitro* and *in vivo* and it could be shown that MRG proteins have a role in regulating splicing variation and flowering (Bu et al., 2014; Xu et al., 2014). Affinity purification analyses should shed light on a possible connection between MRG proteins and splicing. Surprisingly AP-MS experiments and Y2H revealed that MRG proteins may not interact with PTB2 in *Arabidopsis* cells. This indicates that if MRG proteins are involved in the regulation of splicing, the mechanism might be different compared to the situation described in human cells. An impact of MRG proteins on splicing was indicated by AP-MS experiments since a variety of splicing related proteins could be co-purified with MRG proteins. Thus it can be speculated that in *Arabidopsis* MRG proteins are involved in the regulation of splicing in a PTB2 independent way. A connection with splicing factors was more pronounced in MRG2 since more splicing factors were identified as putative interactors in the GS-MRG2 AP-MS experiments than in the GS-MRG1 analyses. Additionally the interactomes of GS-MRG1 and GS-MRG2 were quite different what could be explained by the fact that both proteins share only ~50% sequence identity. These discrepancies could be explained on the one hand if both proteins influence splicing in different ways but it is also possible that both proteins collaborate and one of the two proteins is more involved in the recruitment of splicing factors or simply both proteins may have different functions. Human MRG domains form homodimers and interact with other proteins to form functional complexes (Xie et al., 2015; Zhang et al., 2006). A putative interaction of both MRG proteins was indicated since MRG2 was also identified in the MRG1-SG pull down experiments but this connection was not detected when MRG2 was used as the bait protein. Since it could be shown that *MRG1* and *MRG2* function redundantly (Xu et al., 2014) and dimerization of plant MRG proteins could be observed *in vitro* (Liu et al., 2016) it is possible that both MRG proteins influence splicing in a concerted manner. In upcoming experiments it might be interesting to analyse a possible interaction between the two MRG proteins and how they function in more detail.

Among the putative splicing related MRG interactors several SR proteins were identified. These factors play a key role in plant pre-mRNA splicing and a SR dependent plant-specific mechanism to recognise sequence motives located in exons or introns exists (Jeong, 2017; Shen and Green, 2004). These RNA binding proteins are extensively linked with alternative splicing (Reddy and Shad Ali, 2011) and some SR proteins also interact with H3 tails and associate with the chromatin (Loomis et al., 2009). Additionally SR proteins can directly or indirectly interact with C-terminal domain of RNAPII (De La Mata and Kornblihtt, 2006; Sapra et al., 2009). The finding that SR proteins and MRGs are putative interactors supports the idea that MRG proteins play a role in the recruitment of splicing factors through the chromatin-adaptor model.

Furthermore, a large number of the uridine (U)-rich small nuclear RNPs (snRNPs) especially the pre-catalytic part U5, U4 and U6 but also 17S U2 and U1 were found to co-purify with MRG1 and MRG2. These spliceosomal components are mainly involved in the removal of introns with canonical splice sites (GT-AG).

Additionally, a variety of NTC-associated proteins were identified as putative MRG interactors. In human cells this conserved complex associates with the spliceosome and it is involved in both steps

of intron removal (Hogg et al., 2010; Jia et al., 2017). Interestingly, with many other proteins, SNW/Ski-interacting protein (SKIP), a spliceosome component, was identified when benzonase was not used in the MRG2 pull down, indicating that this interaction may occur due to the nucleic acid binding. The SKIP protein functions as an alternative splicing factor regulating circadian clock and salt tolerance genes (Feng et al., 2015; Wang et al., 2012). SKIP also interacts with ELF7, a component of the conserved Paf1 complex (Paf1c) (Cao et al., 2015; He et al., 2004). Other component of the PAF complex (CTR9, ELF8, CDC73, PAF1) were also found in this study to co-purify with MRG proteins. Since Paf1c promotes specific histone modifications (i.e., H3K4me, H3K36me, and H2Bubi) a connection between factors promoting histone modifications and MRG histone mark readers was disclosed (Tomson and Arndt, 2013).

In summary the co-purification of a variety of different splicing factors suggests that *Arabidopsis* MRG proteins may function as splicing factor recruiters. MRGs possibly interact with spliceosome components and SR proteins and by that influence splicing in a PTB2 independent way.

## 5.2 MRGs co-purify with the methyltransferase SDG7

After the analysis of the AP-MS data indicated that MRG proteins co-purify with splicing factor and are involved in the regulation of splicing, it was further focused on the mechanism how MRGs might be recruited to the chromatin. In the human chromatin adapter complex the MRG15 chromatin reader protein is recruited to the chromatin by H3K36me3 histone marks (Carrozza et al., 2005; Sanidas et al., 2014; Lee and Blenis, 2014). In yeast and humans it is well established that the methyltransferase SETD2 (Set2 in yeast) is mainly responsible for the transcription dependent trimethylation of H3K36 (Chu et al., 2007; Lee and Shilatifard, 2007; Wagner and Carpenter, 2012).

In *Arabidopsis*, 43 proteins contain a conserved SET domain and SDG8 is most similar to the human SETD2 (Xu et al., 2007). A proposed SDG8 - MRG adapter complex could also be described in plants regulating temperature induced AS and flowering time thorough the histone modification, H3K36me3 (Bu et al., 2014; Pajoro et al., 2017; Xu et al., 2014). Pull-down assays and *in situ* PLA showed in *vitro* and in *vivo* the direct interaction of MRG1 and MRG2 with H3K36me3 (Xu et al., 2014). Mutation of SDG8 leads to almost complete abolishment of global H3K36me3 levels and the function of MRG1 and MRG2 in flowering time control is dependent on SDG8-mediated H3K36me3 (Xu et al., 2014) A reduced response to temperature in the *mrg1* and *mrg2* mutants further suggests that H3K36me3 levels affect the splicing outcome (Pajoro et al., 2017).

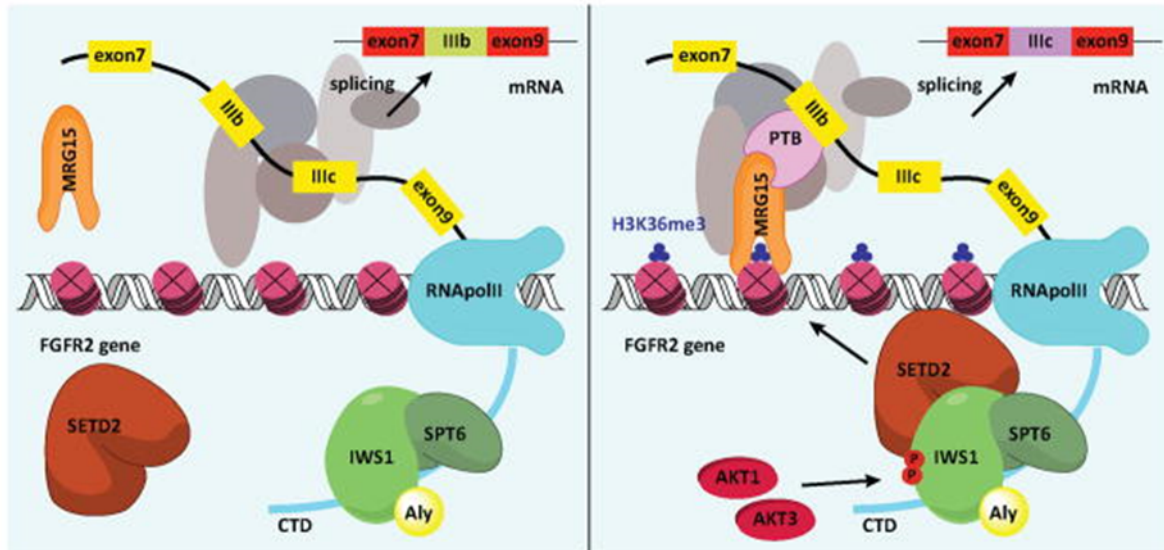
Thus it was further focused if methyltransferases could be identified as putative interaction partners of MRG proteins in the course of the AP-MS data analysis. Two methyltransferases were identified to co-purify with MRG1 and MRG2 but surprisingly SDG8 was not found. One of the two methyltransferases could be identified as SDG7, a close homologue of SDG8. SDG7 co-purified reproducibly with MRG1-SG and GS-MRG2 and the detection was supported with high MASCOT scores. SDG7 was shown to be a negative regulator of vernalization and it is required for proper timing of *VIN3* induction, but the biochemical role of SDG7 in the vernalization response is not clear (Lee et al., 2015). In order to be involved in the deposition of H3K36 histone marks, SDG7 has to be a nuclear protein but SDG7 does not display a predicted nuclear localization signal and SDG7-GFP fusion proteins localize to the cytosol (Lee et al., 2015; Sahr et al., 2010). Lee et al. further showed that the loss of SDG7 has no effect on the global levels of H3K36me3 modifications and that SDG7 shows no histone methyltransferase activity toward individual histones in an *in vitro* assay.

In a previous publication it was shown that SDG7 does not affect global histone methylation (Lee et al., 2015). Another study on the other hand indicated that SDG7 is involved in H3K36 methylation, albeit to a lesser extent than SDG8. Only the loss of SDG7 and SDG8 together fully abolishes both the di- and tri-methylation marks at H3K36 globally, showing that SDG7 and SDG8 may function redundantly (Xu et al., 2014).

The experiments performed in this study to confirm a protein-protein interaction between SDG7 and MRG proteins further suggest that SDG7 might be involved in the H3K36me3 deposition. In contrast to the study performed by Sahr et al., a clear nuclear signal could be observed by CLSM in this study when the localization of SDG7-mCherry fusion proteins was analysed. Additionally a protein-protein interaction between the nuclear MRG proteins and SDG7 could be detected by Y2H and FRET experiments. The confirmation of an interaction between MRG proteins and SDG7 showed that besides SDG8 also SDG7 may be involved in the catalysis of the H3K36me3 – MRG adapter complex in *Arabidopsis*.

### **5.3 IWS1 and MRG proteins may act in the same pathway**

In human cells the MRG15 mediated regulation of splicing is initiated by the TEF IWS1. IWS1 is activated by phosphorylation mediated by the kinases AKT1 and AKT3 and is recruited by SPT6 to the elongating RNAPII complex (Figure 43) (Yoh et al., 2008, 2010, Sanidas et al., 2014). The results obtained in this study so far indicated that MRG1 and MRG2 act as adapters in a chromatin adapter complex and that both factors may be involved in splicing by recruiting several splicing related proteins. Additionally, it could be demonstrated that besides SDG8 (Xu et al., 2014) also SDG7 may act as a 'writer' of the H3K36me3 histone marks which are recognized by the MRG 'reader' proteins. In a next step it was analysed if similar to the situation in human cells also in *Arabidopsis* the TEF IWS1 initiates a putative H3K36me3 - MRG adapter complex, connecting transcript elongation with the co-transcriptional process of splicing.



**Figure 39. Splicing regulation through the IWS1-SETD2-MRG15-PTB2 adapter complex in human cells.** Iws1 is phosphorylated and activated by the kinases AKT1 and AKT3. Activated IWS1 is recruited to the elongating RNAPII by the TEF SPT6 and recruits the methyltransferase SETD2 to methylate H3 at K36 during transcription elongation. These histone modifications are read by the chromatin reader protein MRG15 that recruit the splicing factor PTB2. The figure is taken from (Sanidas et al., 2014).

To analyse if IWS1 and MRG proteins function in the same genetic pathway, the triple mutant *mrg1/mrg2/iws1* was created in the course of this study. Phenotypic analysis of *mrg1/mrg2* mutant plants showed a late flowering phenotype under LD condition what was also observed in a previous study (Bu et al., 2014).

When looking at the time of bolting (Elongation of the first internode), *mrg1/mrg2/iws1* triple mutant plants showed the same late flowering phenotype than *mrg1/mrg2* double mutant plants what would show an epistatic genetic interaction between IWS1 and MRGs. Analysis of other traits on the other hand indicated that IWS1 and MRGs might be involved in the same regulatory pathway. Regarding numbers of leaves and rosette diameter the triple mutant plants showed a statistically significant intermediate phenotype between *iws1* and *mrg1/mrg2* mutant plants. Additionally, a synergistic effect of IWS1 and MRG mutations on primary inflorescences could be observed. Thus although phenotyping did not point to one direction it still indicated that IWS1 and MRGs may act in the same pathway.

## 5.4 IWS1 affects global H3K36me3 levels

In a next step to characterize a putative IWS1-dependent H3K36me3 – MRG adapter model it was tested how the loss of IWS1 and MRGs affects global H3K36me3 levels. Western blotting showed that the loss of IWS1 results in a global reduction of H3K36me3 levels but the loss of both MRGs does not. In the *iws1* and *mrg1/mrg2/iws1* mutant plants a reduced H3K36me3 signal was detected compared to Col-0 and *mrg1/mrg2*, respectively. Thus Western blotting indicated that IWS plays a key role in the regulation of the global H3K36me3 histone mark deposition.

To further analyse the global impact of TEFs on the deposition of H3K36me3 and to address the question how this histone mark might be involved in the regulation of AS in *Arabidopsis*, ChIPseq analysis in *iws1* and *spt4R3* mutant plants using H3K36me3 antibodies was performed. The influence of SPT4 on H3K36me3 levels was also analysed by ChIPseq since in yeast it could be shown that this TEF is also involved in the regulation of the H3K36me3 mark deposition (Lee et al., 2017).

The ChIPseq profile in Col-0 showed that the H3K36me3 marks show the highest density directly downstream the TSS, a feature that could also be observed in a previous publication where H3K36me3 marks were prevalently found at the beginning of the gene body, with a peak at 0.5 kb after the TSS (Pajoro et al., 2017). Notably the reduced level of H3K36me3 in *iws1* observed by western blotting could be confirmed by ChIP-qPCR and ChIPseq. In both mutant lines H3K36me3 levels were clearly reduced downstream the TSS. Also in yeast it could be observed that in the absence of the ATP-dependent chromatin remodel Chd1, which is important to maintain H3K4me3/H3K36me3 domains, aberrant histone H3K36me3 patterns were predominantly observed within 1 kb from the transcription start site (Lee et al., 2017). Strikingly this pattern could also be detected by qPCR after ChIP along the differentially alternatively spliced gene At1g68490. In both mutant lines in the region 1 kb downstream the TSS a clear reduction of H3K36me3 marks could be observed whereas no differences in H3K36me3 occupancies between Col-0 and mutants were detected in more downstream regions. These obtained results indicate that the TEFs SPT4 and IWS1 are both involved in the recruitment of factors which maintain the H3K36me3 histone marks.

Striking differences could be observed when analysing the H3K36me3 levels along exons and introns. In other eukaryotes exons are preferentially marked with H3K36me3 histone marks relative to introns. This H3K36me3 exon marking is dependent on transcription and its level is lower in alternatively spliced exons, supporting a splicing related marking mechanism (Kolasinska-Zwierz et al., 2009). In both mutant lines H3K36me3 exon marks were clearly reduced. Additionally, an enrichment of H3K36me3 at intron/exon boundaries could only be observed in Col-0 but not in *iws1* or *spt4R3*. Thus although reduction of H3K36me3 levels is mainly restricted to a region immediately downstream the TSS, intron and exon definition throughout the gene body seems to be important for H3K36me3 deposition. It can be speculated that H3K36me3 marks at intron/exon boundaries and/or along exons may be involved in the recruitment of the splicing machinery and that in the *tef* mutants this recruitment is disturbed resulting in altered AS pattern.

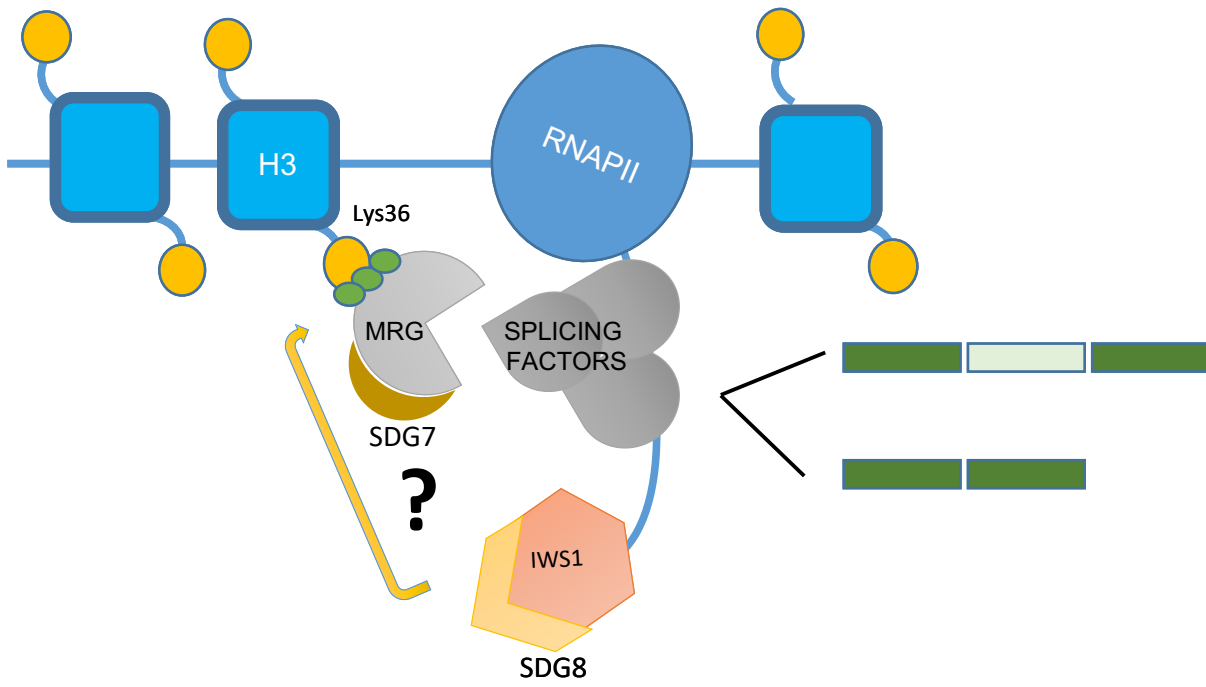
Although it could be shown that H3K36me3 marks are involved in the regulation of splicing, the mode of action still remains to be elucidated (Pajoro et al., 2017). Pajoro et al. could demonstrate that temperature-induced differentially spliced genes are enriched in H3K36me3 and that the reduction of H3K36me3 deposition causes alteration in temperature-induced AS. The preferential location of H3K36me3 mark in the proximity of the AS event could point out an effect on transcription speed in support of the kinetic model. The question arises whether H3K36me3 mediated temperature-induced AS is the result of a slower RNAPII, supporting the kinetic model, or if H3K36me3 histone marks

may recruit splicing factors through the H3K36me3 – MRG adapter complex. More likely both model are involved. Since a reduced response to temperature in flowering induction could be observed in the *mrg1/mrg2* double mutant, it can be supposed that H3K36me3 marks regulate splicing at least partially by influencing the recruitment of splicing regulators via the H3K36me3 – MRG adapter complex (Pajoro et al., 2017).

## 5.5 The loss of IWS1 may result in less binding of MRG2 to the chromatin

To analyse the global impact of a putative IWS1 – H3K36me3 – MRG adapter complex, ChIPseq experiments immunoprecipitating MRG proteins in *iws1* mutant plants should be performed. Additionally, RNAseq in the *mrg1/mrg2* double mutant might shed light on the global impact of MRG proteins on alternative splicing. ChIP experiments conducted in this study could indicate that a reduced H3K36me3 level along a gene, differentially alternatively spliced in *iws1*, correlates with a reduced binding of MRG2 to this region. In the course of this study homozygous mutant lines expressing *eGFP-MRG1* and *eGFP-MRG2* transgenes, with native promoter according the paper of Xu et al., 2014, in Col-0 and the homozygous *iws1* background were generated. Due to time constraints, final ChIPseq experiments could not be performed in this study but these experiments will shed more light on the biological significance of a putative IWS1-dependent H3K36me3 – MRG adapter complex.

The model below proposes how the H3K36me3 – MRG adaptor may affect AS in *Arabidopsis*. The TEF IWS1 has been show with Y2H and BiFC experiments to directly interact with SDG8 in *Arabidopsis* (Wang et al., 2014). Furthermore, IWS1 is involved in the recruitment of the methyltransferase SDG8 to chromatin. SDG8 (perhaps in cooperation with SDG7) deposits the H3K36me3 marks mainly onto regions downstream the TSS but also along exons. By that the H3K36me3 marks either affect the splicing decision by regulating the RNAPII elongation speed, supporting the kinetic model, or by recruiting MRG proteins which in turn recruit splicing factors.



**Figure 40.** Model how the epigenetic mark H3K36me3 may affect the alternative splicing decision in *Arabidopsis*. The TEF IWS1 interacts and activates SDG8 that tri-methylates k36 of H3. MRG proteins bind the H3K36me3mark and recruit splicing factors. SDG7 directly interacts with MRG and might cooperate with SDG8 in the methylation of H3K36me3.



## 6 Materials

### 6.1 Instruments

**Table 7. Kits and instruments used in this study.**

Instrument	Manufacturer /model
2100 Bioanalyzer	Agilent
Blotting	System Semi-dry Blotter Maxi (Carl Roth)
Centrifuges	Sorvall™ Evolution RC and Sorvall™ LYNX 4000 equipped with SLA1500 or SS34 rotors (Thermo Fisher Scientific), Centrifuge 5417R and 5804 R (Eppendorf)
Digital camera	EOS 600D equipped with Macro lens EF-S 60 mm 1:2.8 USM or ETS 18-55 mm objective (Canon)
Homogenizer	TissueLyser II (Quiagen)
Hybridization Oven	Hybridisierungsöfen (Uniequip)
Imager	BioDocAnalyze System (Biometra), Multimage TM FluorChem FC2 (Alpha Innotech) Typhoon FLA 9500 (GE Healthcare)
Library Prep Kit for Illumina	New England Biolabs
Microscopes	TCS SP8 (Leica), ApoTome.2 with Axiocam 503 (Zeiss) SMZ645 stereo microscope (Nikon) with KL 1500 LCD (Schott) Discovery V8 stereo with Axiocam MRc5 and KL1500 LCD (Zeiss)
Plant incubator	Plant incubator (Percival Scientific)
qPCR cyclers	Mastercycler® ep RealPlex (Eppendorf)
Quantum Meter	Quantum Flux ML-200 (Apogee Instruments)
Qubit	Thermo Fisher
Shaking Incubator	Multitron Standart (Infors HT)
Sonicator	Bioruptor® Pico (Diagenode), UW2070 MS73 (Bandelin electronic)
Spektrophotometer	NanoDrop ND-1000 (Peqlab)
Thermocycler	T3000 and T Gradient thermocyclers (Biometra)

### 6.2 Chemicals, solutions and enzymes

Laboratory grade chemicals and reagents were purchased from Applichem (Germany), Carl Roth (Germany), Clonotech, Duchefa (Netherlands), Fluka (Switzerland), Life Technologies (UK), Merck (Germany), Sigma Aldrich (Germany), USBiological (USA), and VWR (USA). Enzymes were purchased from Thermo Fisher Scientific (USA), PEQLAB/VWR (USA) and New England Biolabs (USA). Beads and chemicals for Illumina library preparation were purchased from New England Biolabs.

## 6.3 Oligonucleotides

**Table 8. List of oligonucleotides for RT-PCR, qPCR and genotyping.** Oligonucleotids were obtained from MWG eurofins genomics

Sequence 5' - 3'	Description	Lab ID	target gene
GTTGCCCGTCTCACTGGTGA	Genotyping, T-DNA insertion SALK LBb1.3	812	T-DNA insertion SALK LBb1.3
ATATTGACCATCATACTCATTGC	GABI-KAT LB 8409	1595	LB of GABI-KAT T-DNA
TAGCAGGCTTCCCCTGTCT	lws1a genotyping	1629	At1G32130
TCATGGGCAGATTCAATCTCC	lws1a genotyping	2090	At1G32130
CGTACAACCGGTATTGTGC	Actin 2 fw for genotyping	3384	
GTCAGCAAGGTCAAGACG	Actin 2 rv for genotyping	3385	
GGTCCTCGCGTCTACGGAGC	MRG1 fwd for genotyping	3669	AT4G37280
CCG AAG CTC GAC CTT TTG AAC	MRG1 rv for genotyping	3670	AT4G37280
TACTTTGACAAAGCTTTACCG	MRG2 fwd for genotyping	3671	AT1G02740
CTTTCTTCTCCATTCTTCCA	MRG2 rv for genotyping	3672	AT1G02740
GTGAATCCAAAGCAGAGACTGAT	RT-PCR for alternative splicing analysis	3857	AT1G68490
GCTATCCCTGTCAAGACAATC	RT-PCR for alternative splicing analysis	3858	AT1G68490
GCCGAGAGACGAAAGAGGAAA	RT-PCR for alternative splicing analysis	3859	AT1G69840
GAATACAACCCAAAGCTTGACCC	RT-PCR for alternative splicing analysis	3860	AT1G69840
GTCAAGAGGGCAATGAATGAG	RT-PCR for alternative splicing analysis	3861	AT1G69840
TCGCTGCCTCTCTATTCTA	RT-PCR for alternative splicing analysis	3862	AT1G69840
CAGTGTTGTCGGCAATTTTC	RT-PCR for alternative splicing analysis	3863	AT3G61860
CATTCAACTGATAACCTGCGTT	RT-PCR for alternative splicing analysis	3864	AT3G61860
TGGTGTGAAGTTGTTGTTGC	RT-PCR for alternative splicing analysis	3865	AT3G58510
CTAATCTGTGGCTCAAACCC	RT-PCR for alternative splicing analysis	3866	AT3G58510
TTTGAGCAGAAGTTTCATATTTTC	RT-PCR for alternative splicing analysis	3867	AT1G76180
CATCCATATCACTTAATCAAATGAAA	RT-PCR for alternative splicing analysis	3868	AT1G76180
GGAAAGTTTTAAGGACGACCCC	RT-PCR for alternative splicing analysis	3869	AT3G49430
AAT GTC ACC GGG CAA GTT ACC	RT-PCR for alternative splicing analysis	3870	AT3G49430
GCA ATG TCG AGA TCA AAG TCC AG	RT-PCR for alternative splicing analysis	3871	AT3G49430
CTC AGA GGT ATT GTG ATG GTC AC	RT-PCR for alternative splicing analysis	3872	AT3G49430
CACGTCGTATTGGTCTCCGTAA	RT-PCR for alternative splicing analysis	4038	AT1G68490.2
CCCAGAAAATTGCTCCGTTACC	RT-PCR for alternative splicing analysis	4039	AT1G68490.2
TTGAGCTTGACATGGTGGTC	RT-PCR for alternative splicing analysis	4040	AT3G49430.2
ATAAACCTCTAGCCCAGGGGT	RT-PCR for alternative splicing analysis	4041	AT3G49430.2
GAGACGAAAGAGGAAAAATCCCAA	RT-PCR for alternative splicing analysis	4042	AT1G69840.2
AGCGAACATCGAGCTGTTGA	RT-PCR for alternative splicing analysis	4043	AT1G69840.2
GAGTGGGCCAATTCTCTAACCA	qPCR for chip quality control analysis	4329	AT1G68490

**Table 8.1 (Continuation)** List of oligonucleotides for RT-PCR, qPCR and genotyping. Oligonucleotides were obtained from MWG eurofins genomics

Sequence 5' - 3'	Description	Lab ID	target gene
TAA TGG GTT AGC GAC TCT GCT C	qPCR for chip quality control analysis	4330	AT1G68490
AGT CAT CAA TTG GAG CTT TGT GA	qPCR for chip quality control analysis	4331	AT1G68490
ACGAAACCGACTTGCTACTG	qPCR for chip quality control analysis	4332	AT1G68490
ACC ATC ACC ATC CTT CTC GA	qPCR for chip quality control analysis	4333	AT1G68490
CAG TCT CTG CTT TGG ATT CAC A	qPCR for chip quality control analysis	4334	AT1G68490
TCT CCG TAA CAA CCA TCA CCA	qPCR for chip quality control analysis	4335	AT1G68490
AGCTCCAATTGATGACTGAAACA	qPCR for chip quality control analysis	4336	AT1G68490
AGCATCGATTGTGACCACA	RT-PCR for alternative splicing analysis	4391	AT3g58510
AGC AAA TCA TTT AAC CGG CCA	RT-PCR for alternative splicing analysis	4392	AT3G58510
GGT GTG AAG GTT GTT GTT GC	RT-PCR for alternative splicing analysis	4393	AT3G58510
TCCAAACTCCAATGACAAGCT	RT-PCR for alternative splicing analysis	4394	AT3G58510
TCA ATT GGA GCT TTG TGA ATC C	qPCR for chip quality control analysis	4395	AT1G684990
CGA AAC CGA CTT GCT ACT GAT T	qPCR for chip quality control analysis	4396	AT1g68490
AAA ATC AGT AGC AAG TCG GTT TC	qPCR for chip quality control analysis	4397	AT1g68490
TGG TTA GAG AAT TGG CCC ACT	qPCR for chip quality control analysis	4398	AT1g68490
AGT GGG CCA ATT CTC TAA CCA	qPCR for chip quality control analysis	4399	AT1g68490
TGT TCT GCA CCA TAA CCA TCC	qPCR for chip quality control analysis	4400	AT1g68490
TGT TGT GTG GAA TCT TTG AGG GG	qPCR for chip quality control analysis	4401	AT1g68490
CCC AGA AAA TTG CTC CGT TAC CT	qPCR for chip quality control analysis	4402	AT1g68490
CTC ACC TCT CTT AAT CTC GAT GGA TCG	RT-PCR for alternative splicing analysis	4496	AT1G05090
CGT TTC CTT TAG TGG CTC TAC CAT CTC	RT-PCR for alternative splicing analysis	4497	AT1G05090
TGC TGG ACG AGA TGC GCA GGA A	RT-PCR for alternative splicing analysis	4498	AT1G05090
GAA CTA GTT CTC TGA GCA TCT TTA CCC	RT-PCR for alternative splicing analysis	4499	AT1G05090
GTC ATA CCT TCT GGA ACC TCA ACC T	RT-PCR for alternative splicing analysis	4500	AT1G05090
CCGCAACATCACTCATTGACTTGGG	RT-PCR for alternative splicing analysis	4501	AT1G05090
CGAAGTCAAGGTCACCGTCTCCA	RT-PCR for alternative splicing analysis	4502	AT1G02840
GAAGGTAGAGGAGATCTTGATCTTG	RT-PCR for alternative splicing analysis	4503	AT2G25970
GGACAAGCAGGGGCTGATCAATTTG	RT-PCR for alternative splicing analysis	4504	AT2G25970
GAATTCTAGCTCCAGTCTTAGCTTGC	RT-PCR for alternative splicing analysis	4505	AT2G25970
GGGATAACCGAACAGATTGAACATGC	RT-PCR for alternative splicing analysis	4506	AT2G25970
CCACCCATTGCTGAGTTTCTCATAC	RT-PCR for alternative splicing analysis	4507	AT1G19400
AGAAAGACACCTCCTATGCACTTGA	RT-PCR for alternative splicing analysis	4508	AT1G19400

**Table 8.2 (Continuation)** List of oligonucleotides for RT-PCR, qPCR and genotyping. Oligonucleotides were obtained from MWG eurofins genomics

Sequence 5' - 3'	Description	Lab ID	target gene
CACCTCACTTCTCCAATACATGTCTCC	RT-PCR for alternative splicing analysis	4509	AT1G21970
GGATTTGAAACCATCTCCGGTAGA	RT-PCR for alternative splicing analysis	4510	AT1G21970
ACAACGAGTTCCCCGTAACAATCTC	RT-PCR for alternative splicing analysis	4511	AT1G05360
GCTTCTTGACTCGTTTATCCTCTAC	RT-PCR for alternative splicing analysis	4514	AT2G22090
CCATTTATCTACACATGACGGAGCTAC	RT-PCR for alternative splicing analysis	4515	AT2G22091
GTTAGACTGGGAAGAGCGTTACCTT	RT-PCR for alternative splicing analysis	4549	AT4G09140
CTCTTTTCCAAGTCAACAGCATGC	RT-PCR for alternative splicing analysis	4550	AT4G09140
GCATGCTGTTGACTTGGAAAAGAG	RT-PCR for alternative splicing analysis	4551	AT4G09140
TTTAAAGCTGCTACGCGCTGCA	RT-PCR for alternative splicing analysis	4552	AT4G09140
TCTTGAATCGTCGCAGTCTC	RT-PCR for alternative splicing analysis	4553	AT3G22210
TCCGAGTCGTACACATCTTAAAACG	RT-PCR for alternative splicing analysis	4554	AT3G22210
GTTTCAGTCCAACCTGCTGTGATAGT	RT-PCR for alternative splicing analysis	4555	AT2G25970
GTTAGACTGGAAAGAGCGTTACCTT	RT-PCR for alternative splicing analysis	4556	AT4G38280
CTCTTTTCCAAGTCAACAGCATGC	RT-PCR for alternative splicing analysis	4557	AT4G38280
GCATGCTGTTGACTTGGAAAAGAG	RT-PCR for alternative splicing analysis	4558	AT4G38280
TTTAAAGCTGCTACGCGCTGCA	RT-PCR for alternative splicing analysis	4559	AT4G38280
CCAAAGCAACCGAAAAGTGTT	RT-PCR for alternative splicing analysis	4560	AT5G17320
TCATCCTGAGGCAGATAAACCCCT	RT-PCR for alternative splicing analysis	4561	AT5G17320
CGAAACCGCTGAGACAAATGATG	RT-PCR for alternative splicing analysis	4562	AT2G04039
GGAAAAAGGCCCAAAGGAATATGG	RT-PCR for alternative splicing analysis	4563	AT2G04039
CCATATTCTTTTGGGCGTTTTTCC	RT-PCR for alternative splicing analysis	4564	AT2G04039
ACCTGACCCGTGTGAGGAAATTG	RT-PCR for alternative splicing analysis	4565	AT2G04039
GAG CTA GAG TCT CAA TGC AGA TG	qPCR for chip quality control analysis	4736	AT3G58510
GTA GCA CTG AAC AGC ATT GTC TGT C	qPCR for chip quality control analysis	4737	AT3G58510
AGA TGA TAC AAG CGA TAC GAC AAT G	qPCR for chip quality control analysis	4738	AT3G58510
TGT TGT GTG GAA TCT TTG AGG GG	qPCR for chip quality control analysis	4766	AT1G68490
CCC AGA AAA TTG CTC CGT TAC CT	qPCR for chip quality control analysis	4767	AT1G68490
GTT GTG GTT ATA GGA TGG TTA TG	qPCR for chip quality control analysis	4768	AT1G68490
TGT TCT GCA CCA TAA CCA TCC	qPCR for chip quality control analysis	4769	AT1G68490

**Table 9.** List of oligonucleotides for cloning. Oligonucleotides were obtained from MWG eurofins genomics

Sequence 5' - 3'	Description	Lab ID	target gene
TCC CCCGGG T ATGCTTCTG TTTCA	amplifying cds PTB2 for Y2H insertion in (p69)	SmaI	4001
CG GGATTC CTAATGCATTGGACCAG	amplifying cds PTB2 for Y2H insertion in (p69)	BamHI	4002
TCC CCCGGG T ATGGGAAGCTCGTCGAAG	amplifying cds MRG1 for Y2H insertion in (p69)	SmaI	4003
CG GGATTC TCAGTCGCTTTTCTTTGCC	amplifying cds MRG1 for Y2H insertion in (p69)	BamHI	4004
TCC CCCGGG T ATGGGAAGCCCTAACGCCGCC	amplifying cds MRG2 for Y2H insertion in (p69)	SmaI	4005
CG GGATTC CTA ACC TTC TTT CTT CTC CAT TTC TTC CAC	amplifying cds MRG2 for Y2H insertion in (p69)	BamHI	4006
CG GGATCC ATGGGAAGCCCT	insertion MRG2 in pGreen 5'Gs (p1199)	BamHI	4066
CG GAATTC CTAACCTTCTTTCTT	insertion MRG2 in pGreen 5'Gs (p1199)	EcoRI	4067
CG GGATCC ATGGGAAGCTCG	amplifying cds MRG1 for Y2H insertion in (p68)	BamHI	4068
TCC CCCGGG TCAGTCGCTTTTCC	amplifying cds MRG1 for Y2H insertion in (p68)	SmaI	4069
GCC CGG GCA TGG GAA GCT CGT CGA AG	insertion MRG1 in pCambia2300:p35S::5'eGFP (p810)	XmaI	4317
CGT GAG CTC TCA GTC GTC TTT TCC TTT GCC	insertion MRG1 in pCambia2300:p35S::5'eGFP (p810)	SacI	4318
ACC GGT ACC AAT CAA TTT TTT TAT ATT TAT TTG	cloning MRG1 n. promoter in pCambia2300:p35S::5'eGFP (p810)	kpnI	4420
GTT AAG CTT TGT CAC CGA AAG CGA GAG CGA GA	cloning MRG1 n. promoter in pCambia2300:p35S::5'eGFP (p810)	Hind I	4421
CGT GGA TCC ATG GGA AGC TCG TCG A	cloning MRG1 genomic in pGreen 5'GFP	BamHI	4422
CCG GAG CTC TCA GTC GTC TTT TCC TTT G	cloning MRG1 genomic in pGreen 5'GFP	SacI	4423
CGTGATCCATGGGAAGCCCTAACG	cloning n.p MRG2 in GFP plasmid (p810)	BamHI	4574
CCGCCCGGGTGTAAAGAGACCATCAATAAG	cloning n.p MRG2 in pCambia2300:p35S::5'eGFP (p810)	XmaI	4575
GGGTTGGAACAAAAGGTTAGAGC	amplify genomic MRG2 and cloning in pCambia2300:p35S::5'eGFP (p810)		4590
GCGGAAGCAAGATTCTGTTGATACG	amplify genomic MRG2 and cloning in pCambia2300:p35S::5'eGFP (p810)		4591
GCGGGTACCGCAATGTTCTCTAAACAAAAGTAGT	amplify n.p. IWS1 and cloning in pGreen 3'GFP (p1199)		4610
GCGAAGCTTGGGAAGGGAAGGGAAGGAAG	amplify n.p. IWS1 and cloning in pGreen 3'GFP (p1199)		4611
GCCGTCGACATGGGTTTCGAGGATGAT	amplify genomic IWS1 and cloning in pGreen 3'GFP (p1199)		4612
GCGTCTAGAGAGGTACTTGATCATA	amplify genomic IWS1 and cloning in pGreen 3'GFP (p1199)		4613
GCG CAT ATG CCA GCC AGC AAA AAG ATT TCG	cloning cds ashh3 in pGA (p68)	NdeI	4686
GAC GGA TCC TTA GAC AAT CTC CCA GTC TTC TCT CGA C	cloning cds ashh3 in pGB (p69)	BamHI	4687
GCAGAGCTCATGGGAAGCCCT	cloning cds MRG2 in 3'eGFP (p1221) for FRET	SacI	4706
CGC GGA TCC ACC TTC TTT CTT CTC CAT T	cloning MRG2 in 3'eGFP (p1221) for FRET, without stop codon	BamHI	4740
CGG GAT CCC TAA CCT TCT TTC TTC TCC ATT	cloning Mrg2 pCambia2300:p35S::5'eGFP (p810)	BamHI	4833
GCA TCT AGA ATG CCA GCC AGC AAA AAG ATT TCG	cloning ASHH3 in pGreen0179:p35S::mCherry (p1289) for FRET	XbaI	4857
GCA GAG CTC GAC AAT CTC CCA GTC TTC TCT CGA C	cloning ASHH3 in pGreen0179:p35S::mCherry (p1289) for FRET	SacI	4859
TCC GAG CTC ATG GGA AGC TCG	cloning MRG1 in pCambia2300:p35S::5'eGFP (p810) for FRET	SacI	4923
CGG GAT CCT CAG TCG TCT TTT CC	cloning MRG1 in pCambia2300:p35S::5'eGFP (p810) for FRET	BamHI	4924
TCC GTC GAC ATG GGA AGC TCG	cloning cds MRG1 in 3'eGFP (p1221) for FRET	Sall	4925
CGG GAT CCG TCG TCT TTT CCT TT	cloning cds MRG1 in 3'eGFP (p1221) for FRET	BamHI	4926

**Table 10. Bacterial and Yeast strain.**

Organisms	Strain	Resistance	experiment	Company
E. coli	XL1blue	Tetracycline	Plasmid amplification	Stratagene
A. tumefaciens	GV3101:pMP90 + pSOUP	Gentamycin, Tetracycline	Plant transformation	DSMZ S
S. cerevisiae	AH109	-Ade -His -Leu -Trp	Y2H	Clontech

## 6.4 Plasmids

**Table 11. List of vectors created in the course of this study. from this study.** (\* = from this study, A= Ampicillin, Exp. = in which experiment has been used, K = Kanamycin, L = lab collection, No. = lab ID, R. = Plant resistance marker, R. sites = Restriction sites, S= Source).

No.	Name Plasmid	Exp.t	Description	R	R.sites	S
810	pCambia2300-35S:EGFP-modifiedMCS	FRET	3555/3557 (5'eGFP plus MCS) into pCambia2300	K	XbaI/EcoRI	*
853	pGreen0179-pPTB2-PTB2-CDS-SG	AP-MS	#1353 From Andreas Wachter lab	K		L
854	pCambia2300-MRG1-CDS-GS	AP-MS	#1348 from Andreas Wachter lab	K		L
855	pCambia2300-MRG2-CDS-GS	AP-MS	#1354 from Andreas Wachter lab	K		L
1199	pGreen0179:3'GFP	OEX	pGreen0179 #1094 digest with SacIHF+EcoRIHF,insert eGFP cut with same enzyme from pCaqmbia35S-MCS-eGFP( Irene)	K	sacI/EcoRI	L
1205	pGreen0179:n.p.IWS 3'GFP	OEX	pGreen0179 #1199digest withKpnIHF+HindIIIHF,insert n.p IWS cut with same enzyme	K	KpnI/HindIII	*
1228	pGreen0179:n.p.genomic iws1a 3'GFP	OEX	n.p. and genomic #iws1a cloned in 1205	K	Xba /Sall	*
1235	pGADT with CDS ashh3	Y2H	from plasmid 990, cut and clone in with NdeI and BamHI	A	BamHI/NdeI	*
1259	pGreen0229 35s 5'mCherry ashh3	FRET	Backbone from plasmid 1212 insert	K	BamHI/NdeI	*
1269	pGBKT7 MRG1	Y2H	Backbone from plasmid ori68, Insert amplified with primer 4003 and 4004	K	SmaI /BamHI	*
1270	pGBKT7 MRG2	Y2H	Backbone from plasmid ori68, Insert amplified with primer 4005 and 4006	K	SmaI /BamHI	*
1273	pCambia 2300 35s mrg2 3'EGFP	OEX	backbone 1121, cut 1253 and insert with XbaI and BamHI	K	XbaI/BamHI	*
1279	pCambia2300 35S eGFP_MRG2	FRET	backbone 810, insert amplified with primer 4574-4006	K	BamHI/SacI	*
1305	pGBKT7-ASHH3	Y2H	backbone 992 insert 1235	K	NdeI/BamHI	*
1316	pCambia2300 535:ASHH3_3'mcherry	FRET	vector backbone 1289, CDS amplified with primer 4857-4859	K	sacI/XbaI	*
1321	pCambia2300 535:5'EGFP_cdsMRG1 (lost plant resistance)	FRET	backbone1279, insert amplified with primers 4923-4924	K	SacI/BamHI	*
1428	pCambia2300-pPTB2-PTB2-CDS-SG	AP-MS		K		*
1429	pCambia2300-GS-MRG1-CDS	AP-MS		K		*
1430	pCambia2300-GS-MRG2-CDS	AP-MS		K		*

## 6.5 T-DNA and RNAi lines

**Table. 12** List of T-DNA and RNAi lines.

Name	AGI	T-DNA insertion	RNAi line	Source
mrg1-1	At4g37280	SALK_057762		NASC
mrg2-3	At1g02740	GK-255G06		NASC
tflls	At2g38560	SALK_056755		NASC
iws1	At1g32130	SALK_056238		NASC
Spt4R3	At5g63670		RNAi 3	NASC

## 6.6 Databases, Online Tools and Software

**Table 13.** Databases, Online Tools and Software.

Bioanalyzer data analysis	( <a href="https://www.agilent.com/">https://www.agilent.com/</a> )
Geneinvestigator	( <a href="https://geneinvestigator.com/gv/">https://geneinvestigator.com/gv/</a> )
ImageJ 1.49d	( <a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a> )
Integrative Genomics Viewer	( <a href="https://software.broadinstitute.org/software/igv/">https://software.broadinstitute.org/software/igv/</a> )
Leica Application Suite X Leica Microsystems	( <a href="https://www.leica.microsystems.com">https://www.leica.microsystems.com</a> )
Mendeley	( <a href="https://www.mendeley.com/">https://www.mendeley.com/</a> )
Microsoft Excel 2016	( <a href="https://www.microsoft.com/">https://www.microsoft.com/</a> )
OligoCalc	( <a href="http://biotools.nubic.northwestern.edu/OligoCalc.html">http://biotools.nubic.northwestern.edu/OligoCalc.html</a> )
Past3.26	( <a href="http://folk.uio.no/ohammer/past/">http://folk.uio.no/ohammer/past/</a> )
PLAZA 4.0	( <a href="https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/">https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/</a> )
Primer3 v 0.4.0	( <a href="https://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm">https://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm</a> )
Snappgene v2.3.2	( <a href="https://www.snappgene.com/">https://www.snappgene.com/</a> )
The Arabidopsis Information Resource - v10	( <a href="http://www.arabidopsis.org/">http://www.arabidopsis.org/</a> )
UniProt	( <a href="https://www.uniprot.org/">https://www.uniprot.org/</a> )
Venny v2.1	( <a href="https://bioinfogp.cnb.csic.es/tools/venny/">https://bioinfogp.cnb.csic.es/tools/venny/</a> )





## 7 Methods

### 7.1 Nucleic Acid base method

#### 7.1.1 Genomic DNA isolation from *Arabidopsis* leaves

*Arabidopsis* leaves were harvested in a 1.5 mL Eppendorf tube with two-three glass beads inside, frozen immediately in liquid nitrogen and homogenized using the Tissue Lyser II (Qiagen) with a frequency of 30 Hz for 30 secs. The Edwards buffer (200 mM Tris pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5 % (w/v) SDS), (Edwards et al., 1991) was freshly prepared and added to the ground tissue. The sample was vortexed for 5 secs and then centrifuged for 1 min at full speed and room temperature (RT). An equal volume (300  $\mu$ L) of 100 % (v/v) isopropanol were mixed with the supernatant to precipitate the DNA and incubated at RT for 2 min. After centrifugation RT for 5 min at 12000 g the DNA pellet was washed once with 70 % (v/v) ethanol, air dried and resuspended in 50  $\mu$ L H<sub>2</sub>O.

#### 7.1.2 RNA isolation from *Arabidopsis* leaves (for alternative splicing experiment)

14 days old seedlings after stratification, grown on MS plates in the plant incubator, were harvested with tweezers and roots cut with scissors. The areal part was then immediately frozen with liquid nitrogen in a mortar and ground with a pestle. The plant powder was transferred in 1.5 mL tube and stored frozen. RNA was extracted from 50-100mg of tissue sample after the addition of 1mL TRIzol™ reagent (Invitrogen). Homogenized samples were centrifuged at 12000g for 10 minutes, 4° C. The clear supernatant with RNA was transferred to a new tube. The homogenized sample was incubated for 5minutes at RT to permit complete dissociation of the nucleoprotein complex. 0.2mL of chloroform per 1mL of TRIzol reagent was added, the sample shaken for 15 secs by hands, incubated 2-3 minutes at RT. Samples were centrifuged for 15 minutes at 4°C and 12000g. The aqueous phase of the sample was transferred to a new cup. 0.5 mL of 100% isopropanol was added to the aqueous phase, incubated at RT for 10 minutes and centrifuged at 12000g for 10 minutes at 4°C. The pellet was washed with 1 mL of 75% ethanol, vortexed briefly and centrifuged at 7500g for 5 minutes at 4°C. The RNA pellet was vacuum dried for 5-10 minutes. The RNA pellet was suspended in 50  $\mu$ L RNase-free water.

#### 7.1.3 DNase-digest

4  $\mu$ g of isolated RNA were DNaseI-treated in a total volume of 20  $\mu$ l with 2  $\mu$ L DNaseI (200 u/ $\mu$ L) and 2  $\mu$ L DNaseI buffer 10x (Thermo Fisher Scientific), incubated for 90 minutes at 37°C. Then 1  $\mu$ L of 50 mM EDTA was added and incubated for 10 minutes at 65°C.

### 7.1.4 Reverse transcription (cDNA synthesis)

1.5 µg DNase-digested RNA was transcribed into cDNA using RevertAidTMH Minus M-MuLV Reverse Transcriptase (Thermo Fisher Scientific) in a total volume of 11 µL with 0.5 µg oligo-dT primers for 5 min at 70 °C and cooled down to 4 °C. Reaction buffer (1x), dNTP (1mM) and 20 U RNase Inhibitor (Thermo Fisher Scientific) were added to a total volume of 19 µL and the mixture was incubated for 5. min at 37 °C. To synthesize cDNA, 200 U of RevertAidTMH Minus M-MuLV Reverse Transcriptase were added and the sample was incubated for 60 min at 40 °C and then for 10 minutes at 70°C. The reaction was then stopped after 10 minutes at 4°C.

### 7.1.5 Polymerase chain reaction

In general, Taq DNA Polymerase (PEQLAB) was used for any standard PCR. Hercules II Fusion DNA Polymerase and KAPA HiFi (Peqlab) DNA Polymerase were used for cloning due to their proofreading activities. The PCR reaction conditions and cycle programs are shown in Table 14 and Table 15, respectively. Amplified fragments were analysed on 0,7 – 2 % agarose gels depending on the size of the amplified fragment.

**Table 14. PCR cycle programs for Taq, KAPA HiFi and Hercules II.**

A. Settings to amplify genomic or plasmid targets B. Settings to amplify cDNA targets

Step	Taq		KAPA HiFi		Hercules II		Cycle
	Temperature [°C]	Time [sec]	Temperature [°C]	Time [sec]	Temperature [°C]	Time [sec]	
initial denaturation	95	300	98	300	95	120	1
denaturation	95	30	98	20	95	20	1
annealing	primer Tm -5		30 primer Tm -5		primer Tm -5		20 25-35
extension	72	60/1000bp	72	30 / 1000bp	68A / 72B	30 sec. per 1 kbA /60 sec. per 1 kbB	1
final extension	72	300	72	300	68/72	68/72	1
Storage	4 ∞		4 ∞		4 ∞		

**Table 15. Reagents used for Taq, KAPA HiFi and Hercules II.**

Reagent	Taq	KAPA HiFi	Hercules II
Buffer	1x Taq reaction buffer (Peqlab)	1x KAPA HiFi Buffer (Peqlab)	1x Herculanse reaction Buffer (Agilent)
dNTP mix	0.2 mM of each dNTP	0.3 mM of each dNTP	0.3 mM of each dNTP
Forward primer	0.5 µM	0.25 µM	0.2µM
Reverse primer	0.5 µM	0.25 µM	0.2µM
Polymerase	0.5 U Taq DNA polymerase (Peqlab)	1U KAPA HiFi (Peqlab)	0.5 U Herculanse polymerase (Agilent)
MilliQ water	Up to 25µl	Up to 50µl	Up to 50µl

### 7.1.6 Real time quantitative PCR (qRT-PCR)

qPCR reaction was performed in 10  $\mu$ L of total volume using KAPATM SYBR® FAST QPCR Master Mix Universal (PEQLAB), G003-SFstripes (Kisker Biotech GmbH and Co KG) and the Mastercycler epgradient S realplex 2 with realplex software v2.2 (Eppendorf AG) according to the manufacturer's instructions. Targets were amplified with specific primer pairs (Table 9) using the program primer3 (Untergasser et al.,2007). The cycling program in table 16 was used. For ChIP analysis, 2  $\mu$ L of immunoprecipitated DNA (following desired dilution with H<sub>2</sub>O) was analysed with locus specific primers. Data was analysed using the  $2^{-\Delta\Delta CT}$  method and normalized to the input samples.

**Table 16.** qRT-PCR cycle program

Step	Temperature [°C]	Time [sec]	Cycle
initial denaturation	95	180	1
denaturation	95	3	40
annealing	60	20	
extension	72	8	
Melting curve	95	15	1
	60-95 (gradient)		1
	95	15	1

### 7.1.7 Reverse Transcriptase PCR (RT-PCR) for alternative splicing analysis

RT-PCR was performed to amplify different alternative splicing isoforms, using the T3000 thermocycler (Biometra GmbH, Göttingen). Taq DNA Polymerase (PEQLAB) was used according to table 15. Primers were design flanking specific alternative event. Additional Mg<sup>2+</sup> was added to the PCR mix. In the cycle program a Hot Start was applied to reduce the presence of primer dimers. From the final 25  $\mu$ L of PCR product 15  $\mu$ L were removed for the Bioanalyzer quantification analysis and 10  $\mu$ L were used for the agarose gel electrophoresis or polyacrylamide gel electrophoresis.

### 7.1.8 Calculation of Splicing isoforms

Isoform concentrations were measured using the Agilent 2100 Bioanalyzer with the DNA7500 kit. Oligonucleotides used are listed in Table 9. Gel pictures were enhanced using the Adobe Photoshop autocontrast function. 15  $\mu$ L RT-PCR product were vacuum concentrated for 20 minutes. Alternative splice forms were co-amplified and quantified using a Bioanalyzer. Ratios were normalized to the corresponding control Col-0 samples. The quantification analysis was performed according to manufacturer's instructions. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS.

### 7.1.9 Alternative splicing PCR isoforms sequencing

Splice variants were subcloned using the pGEM®-T Vector System I (Promega) and sequenced.

### 7.1.10 Agarose gel electrophoresis

1 - 2% (w/v) agarose gels was boiled in 1x TAE buffer (40 mM Tris pH 8, 20 mM acetic acid, 1 mM EDTA) and supplemented with 0.005 % (v/v) ethidium bromide. DNA/RNA samples were mixed with 6x loading dye (250mM Tris pH 7.5, 10% w/v SDS, 30% v/v glycerol, 0.5 M DTT, 0.1% w/v bromophenol blue) and separated. Gels were run at 150 V. Samples were loaded into the wells of the agarose gel. DNA/RNA fragments were visualized by excitation at 256 nm with a BioDoc Analyser (Biometra GmbH, Göttingen).

### 7.1.11 DNA extraction from Agarose gel and PCR clean up

The NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) was used according to the manufacturer's instructions for the clean-up of PCR samples and DNA fragments from agarose gels.

### 7.1.12 Restriction digest and dephosphorylation

Plasmid and PCR fragments were digested with restriction enzymes from NEB according to the manufacturer's instructions. Digestions were performed o/n. To prevent self-ligation, 5' - phosphate groups from digested plasmids were removed by incubation with 5 U of Antarctic Phosphatase (NEB) in 1 x Antarctic Phosphatase buffer for 60 minutes at 37 °C.

### 7.1.13 Ligation

After digestion, inserts and plasmids were mixed in a 4:1 molar ratio with the addition of 5U T4 DNA ligase (Thermo Scientific) and 1x T4 Ligase buffer in a total volume of 20 µL. Ligation was performed o/n at 4 °C in a thermos canister.

### 7.1.14 Isolation of plasmid DNA from *E. coli*

For minipreparation of plasmid DNA, 2 mL selective LB-medium were inoculated with a positive transformed *E. coli* colony and incubated o/n at 37 °C and 200 rpm. Cells from 1.5 ml liquid cultures were harvested by centrifugation for 3 min at 1800 g. Cell pellet was resuspended in 200 µL P1 buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 100 µg/mL RNase A). For cell lysis, 300 µL of P2 buffer (0.2 M NaOH, 1 % (w/v) SDS) were added and the mixture was incubated for 5 min at RT. To stop the cell lysis, 300 µL of P3 buffer (3 M potassium acetate, pH 4.8) were added and the sample was incubated for 10 min on ice before centrifugation at 12000 g for 10 min and RT. The supernatant was transferred to a new 1.5 mL tube and an equal amount of 100 % (v/v) isopropanol was added.

After incubation for 5 min at RT to precipitate the plasmid DNA, samples were spun down at 12000 g for 10 min at RT and the pellet was washed with 70 % (v/v) ethanol, air dried and re-dissolved in 50  $\mu$ L H<sub>2</sub>O.

Midipreparation of plasmid DNA was performed using the NucleoBond® Xtra Midi Kit (Macherey Nagel) according to the manufacturer's instruction.

### 7.1.15 Sequencing

Sequencing of PCR products or desired fragments of purified plasmid DNA was performed by the TubeSeq Service of Eurofins MWG Operon (Ebersberg). DNA samples and sequencing primers were prepared according to instructions of the provider (<https://www.eurofinsgenomics.eu>).

## 7.2 RNAseq

RNA-seq analysis was performed using mRNA from the four mutant plant lines, *tfl1s*, *iws1*, *tfl1s/iws1*, Spt4R3 and transcriptomes were compared regarding variation in splice variants compared to Col-0 based on three replicates per genotype. To obtain data of high quality that allows rigorous statistical analysis of detected splicing events, we aimed for ~90 Mio reads per sample (Figure 9). Library preparation and RNAseq were carried out according to the NEBNext Ultra™ Directional RNA Library Prep protocol (New England Biolabs, Ipswich, MA, USA), the Illumina HiSeq 1000 System User Guide (Illumina, Inc., San Diego, CA, USA), and the KAPA Library Quantification Kit - Illumina/ABI Prism User Guide (Kapa Biosystems, Inc., Woburn, MA, USA). In brief, 1000 ng of total RNA were used to purify poly-A containing mRNA molecules, using the NEBNext Poly(A) mRNA Magnetic Isolation Module. Next, the mRNA was reverse transcribed into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis, incorporating dUTP residues. The resulting cDNA fragments subsequently went through an end repair process, the addition of a single 'A' base, the ligation of the barcode-containing adapters (NEBNext Multiplex Oligos, New England Biolabs), and a purification step. To preserve the strand orientation, the dUTP containing second strand was excised by the USER enzyme and finally the DNA libraries were amplified by PCR. The libraries were quantified using the KAPA SYBR FAST ABI Prism Library Quantification Kit. Equimolar amounts of each library were pooled, and the pools were used for cluster generation on the cBot (TruSeq PE Cluster Kit v3). The sequencing run was performed on a HiSeq 1000 instrument using the indexed, 100 cycles single-read (SR) protocol and TruSeq SBS v3 Reagents. Image analysis and base calling resulted in .bcl files, which were converted into .fastq files with CASAVA 1.8.2 software.

Library preparation and RNAseq were performed at the Genomics Core Facility "KFB - Center of Excellence for Fluorescent Bioanalytics" (University of Regensburg, Regensburg, Germany; [www.kfb-regensburg.de](http://www.kfb-regensburg.de)).

To identify and quantify splicing events, the resulting reads (100 bp paired-end per genotype) were mapped to the Arabidopsis genome (TAIR10 version) resulting in approximately 130 mio. reads per library (fig.9) based on three replicates per genotype.

## 7.3 ChIP seq

### 7.3.1 Material preparation

14 DAS seedlings grown on sterile plates with MS medium were used for ChIP experiments. 1-1.5 g Plant material was harvested and subjected to DNA crosslinking with 1% formaldehyde. The seedlings were vacuum infiltrated for 10 minutes. Formaldehyde was quenched by adding 2.5 mL of 2.5 M Glycine and again vacuum infiltrated for 10 minutes. Seedlings were then washed and frozen in liquid nitrogen.

#### Chromatin Preparation

Plant material was transferred into the Tissue Lyser II adapters (Qiagene) prechilled in liquid nitrogen and samples were homogenized with the frequency of 30 Hz for 1 min. Ground tissue was added to 30 mL Extraction Buffer 1 in a 50 mL Falcon tube and incubated on a rotating wheel at 4 °C for 10 - 20 min. Solution was filtered through a double layer of Miracloth into a new 50 mL Falcon tube and centrifuged at 3000 g and 4 °C for 20 min. Pellet was resuspended in 1 mL Extraction Buffer 2, transferred into 1.5 mL tube and centrifuged at 12000 g and 4 °C for 10 min. Washing was repeated 3 times. Pellet was next resuspended in 400  $\mu$ L Extraction Buffer 3 and another 400  $\mu$ L of Extraction Buffer 3 was added to a new 1.5 mL tube and overlaid with the pellet from the previous step. Following centrifugation at 16000 g and 4 °C for 1 h pellet was resuspended in 400  $\mu$ L Nuclei Lysis Buffer and incubated on ice for 30 min. Each sample was sonicated using Bioruptor® Pico (Diagenode) with 10 cycles of 30 sec on/30 sec off. 50  $\mu$ L sample were collected before and after the sonification to test it. Chromatin solution was centrifuged twice at 14,000 rpm and 4 °C for 10 min. Supernatant was transferred into new 1.5 mL tube containing agarose beads (Sigma Aldrich) pre-washed in ChIP Dilution Buffer according to manufacturer's manual. The mixture was incubated at 4 °C with agitation for 2 hours to reduce the background signal. Beads were routinely pelleted by centrifugation at 2000 g for 1 min.

#### Immunoprecipitation (IP)

The supernatant was diluted 10x into a new 13 mL tube with ChIP Dilution Buffer. Chromatin solution was divided into several 1.5 tubes with 300-500  $\mu$ L per IP. Desired antibodies were added including  $\alpha$ -H3,  $\alpha$  H3K36me3,  $\alpha$  -CTD-S2P,  $\alpha$  -CTD-S5P (Abcam, 5  $\mu$ g each) or GFP-trap (Chromotek, 30  $\mu$ L). No antibodies were added as a negative control (NoAb). Samples were left rotating overnight at 4 °C.

#### Washing and Elution

Chromatin solution with antibodies was added to Protein A agarose beads (Millipore) pre-washed in ChIP Dilution Buffer (usually 40  $\mu$ L beads per 500  $\mu$ L buffer, 3 times washed, beads pelleted by centrifugation 2000g for 1 minute). Following 3 hours rotation at 4 °C beads were pelleted and sequentially washed with 0.5 mL of washing buffers (buffers 6-9 below). Before the first washing, 500  $\mu$ L of the supernatant were collected for the negative control to determine chromatin concentration (Input). After each washing step samples were left rotating at 4 °C for 10 min. Following the last washing step beads were pelleted by centrifugation and DNA was eluted by adding 250  $\mu$ L freshly made Elution Buffer following supernatant removal. Samples were next incubated at 65 °C for 15 min with gentle rotation and the supernatant above pelleted bead was carefully transferred into a new 1.5 mL tube.

## DNA Extraction

Crosslinking was reversed by adding 4  $\mu$ L of 5 M NaCl to the 100  $\mu$ L of eluate followed by overnight incubation at 65 °C with agitation. Next, 2  $\mu$ L 0.5 M EDTA, 1.5  $\mu$ L 3 M Tris-HCl pH 6.8 and 2 mg proteinase K was added to the eluate and samples were incubated at 45°C for 3 h with agitation. DNA was purified using Phase Lock Gel tubes from Eppendorf. The phase lock gel tubes were first spun down for 30 sec at 16000g. 500  $\mu$ L (equal volume) 25/24/1 phenol/chloroform/isoamyl alcohol were added to the gel tube and mixed. Samples were then poured into the Phase Lock Gel tubes and mixed for 30 secs. Samples were centrifuged to separate phases at 16000g for 5 minutes. 500  $\mu$ L of 1-brom-3-chloropropan were added and centrifuged 5 minutes at 16000g. 400  $\mu$ L were transferred in a fresh tube. 40  $\mu$ L NaAc 3M pH 5.8 and 1 mL EtOH 100% were added and samples were kept for 1h at -80 °C. Samples were then centrifuged at full speed and 4 °C for at least 30 minutes. Supernatant was removed and the pellet was washed with 1 mL 70% EtOH. The Pellet was resuspended in 40  $\mu$ L H<sub>2</sub>O and further dilutions were made for each sample before proceeding with PCR.

**Table 17. Buffers used for ChIP.**

	Extraction 1	Extraction 2	Extraction 3	Nuclei Lysis	ChIP dilution	Low salt	High salt	LiCl	TE	Elution
Sucrose	0.4 M	0.25 M	1.7 M							
HEPES pH 8	10 mM	10 mM	10 mM	50 mM	16.7 mM	20 mM	20 mM			
$\beta$ mercaptoethanol	5 mM	5 mM	5 mM							
Triton X-100		1%	0.15%		1.10%	1%	1%			
MgCl <sub>2</sub>		10 mM	2 mM							
EDTA				10 mM	1.2 mM	2 mM	2 mM	1 mM	1 mM	
SDS				0.10%		0.10%	0.10%			1%
NaCl					167 mM	500 mM	150 mM	50 mM		
LiCl								0.25 M		
sodium deoxy								1%		
NaHCO <sub>3</sub>										0.1 M
NP-40								1%		
Tris-HCl									10 mM	

Buffers 1-5 were additionally supplemented with  $\beta$ -Glycerophosphate (Sigma-Aldrich), cOmplete™ EDTA free proteinase inhibitor tablets (Sigma-Aldrich) and PhosSTOP™ (SigmaAldrich) according to according to the manufacturer's manual.

### 7.3.2 DNA quality control for ChIP seq

Quality check and quantification of DNA samples before proceeding with sequencing analysis was performed. DNA concentrations after immunoprecipitation were measured using a Qubit device and size distributions of adapter-ligated fragments were calculated using a Bioanalyzer (Figure 19, S). DNA fragmentation was achieved mechanically by sonication (Bioruptor® Pico sonication device). Samples were used for sequencing when fragment size ranged between 100 and 350 bp and the DNA concentration was higher than 2nM.

### 7.3.4 Library preparation and sequencing

Adaptor ligation and library amplification steps were done following the standard protocol using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® sequencing. The final sequencing was performed at the Genomics Core Facility “KFB - Center of Excellence for Fluorescent Bioanalytics” (University of Regensburg, Regensburg, Germany; [www.kfb-regensburg.de](http://www.kfb-regensburg.de)) using a NextSeq High-Output Flow Cell (2x75 Paired-end). The ChIP-seq data analysis was performed in our laboratory by Simon Obermeyer.

Obtained reads were aligned to the Arabidopsis reference genome (TAIR10).

Reads were demultiplexed (by KFB) and trimmed with TRIMMOMATIC (Illumina Clip). Duplicate reads were filtered and removed using PICARD and aligned to the TAIR10 genome using BOWTIE. Quality assessment was performed at each step with FASTQC. Alignment files were INPUT normalized, read centered, extended to fragment size and sequencing depth was normalized with bam Compare (deepTools).

## 7.4 ChIP seq data analysis

### RNAPII occupancy and H3K36me3 distribution

The bioinformatics analysis was performed by Simon Obermeyer.

The reads from 2 biological replicates were demultiplexed (by KFB), trimmed with TRIMMOMATIC (Illumina Clip), duplicate reads were filtered and removed using PICARD. Reads were aligned to the TAIR10 genome using BOWTIE. The quality assessment was performed at each step with FASTQC. The alignment files were INPUT normalized, read centered, extended to fragment size and sequencing depth normalized with DAMPOS deep Tools.

ChIP-seq tracks showing RNAPII-S2P, RNAPII-S5P and H3k36me3 reads over single genes were generated using Integrative View Genomics (IGV).

All introns and exons were isolated from the TAIR10 genome annotation and additionally filtered for introns and exons that are larger than 500bp only.

Motif analysis was performed using HOMER v4.11 (<http://homer.ucsd.edu/homer/>).

## 7.5 Protein based methods

### 7.5.1 Protein extraction from Arabidopsis seedlings

14 DAS seedlings were ground using pestle and mortar. 250 µL of Extraction Buffer (10mM MgCl<sub>2</sub>, 50mM Tris HCl pH 7.5, 150 mM NaCl, 0,1% NP40, 1mM PMSF, PI(25x) ) was added to 50 mg ground material. Samples were centrifuged 10 minutes at 13000g and 4°C. The supernatant was transferred into a new tube. Centrifugation was repeated, the supernatant was transferred into new tube and 50 µL, 6X SDS loading dye ( 50 mM Tris pH 6.8, 0,002% (w/v) bromophenol blue, 2,5% (w/v) glycerol, 1 % (w/v) SDS and 143 mM -mercaptoethanol) were added.



## 7.5.2 SDS-PAGE

Different resolving gels were prepared depending on the size of the proteins to be analysed with (9%, 12% or 18% (w/v) acrylamide: bisacrylamide (30:0.15), 0.75 M Tris pH 8.8, 0.2 % (w/v) SDS, 0.1 % (w/v) ammonium persulfate (APS) and 0.02 % (v/v) TEMED)). The stacking gels contained acrylamide: bisacrylamide (30:0.8) (w/v), 0.14 M Tris-HCl pH 6.8, 0.23% SDS (w/v), 0.11% APS (w/v) and 0.06% TEMED (v/v). Gels were cast in a Bio-RAD Mini-Protean® 3 Multicaster system (Bio-Rad).

Analysed samples were mixed with 6 x SDS loading buffer to a final concentration of 1x and heated at 95 °C for 10 min. SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed in a Bio-RAD Mini-Protean® 3 running chamber using Laemmli running buffer (0.1% SDS (w/v), 3.03 g/L Tris, and 14.41 g/L glycine). Gels were run at 160-200 V and the size of proteins were estimated using Page Ruler™ unstained protein ladder (Thermo Scientific).

## 7.5.3 Coomassie Brilliant Blue (CBB) staining

The proteins were visualized after incubating the polyacrylamide gel for 1h at RT in Coomassie Brilliant Blue (CBB) solution (0.2 % (w/v) CBB G-250, 30 % (v/v) ethanol and 10 % (v/v) acetic acid) and subsequently destaining with 7.5 % (v/v) ethanol and 5 % (v/v) acetic acid o/n. Gels were documented with a digital camera EOS 600D (Canon) .

## 7.5.4 Western Blot

Before transferring the proteins from polyacrylamide gel onto a Hybond LFP 0.2 PVD membrane (Amersham), the blotting membrane was activated by rinsing with 100% methanol for 30 sec and subsequently equilibrated together with Whatman paper (Biometra) in blotting buffer (20% (v/v) methanol, 200 mM glycine, 20mM Tris, 0,01% (w/v) SDS). Proteins were transferred in blotting buffer onto the PVDF membrane using a Semidry Blotter Maxi (Roth) at 50 Ma for 3h. Unspecific binding sites were blocked by incubating the membrane for 1h in 15 mL blocking buffer (5 % (w/v) skimmed milk powder, 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % (v/v) Tween 20) on the rotating wheel at 4 °C. The primary antibody was added directly to blotting buffer in a 1:2000 dilution and incubated o/n at 4 °C with agitation.  $\alpha$ GFP (ChromoTeq),  $\alpha$ -CTD-S2P,  $\alpha$ -CTD-S5P, H3, H3K36me3, H3K27me3 (Abcam) antibodies were used.. The membrane was washed three times for 10 min. with washing buffer (0.1 % (v/v) Triton X-100, 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05 % (v/v) Tween 20). Respective secondary antibodies were mixed in 1:5000 dilutions with blotting buffer and incubated with the membrane for 1 - 2 hours at 4 °C with agitation. For chemiluminescent detection,  $\alpha$ -Rabbit,  $\alpha$ -Mouse or  $\alpha$ -Rat IgG antibodies coupled with horseradish peroxidase (HRP) (Sigma Aldrich) were used. For fluorescent detection, 1: 5000 ECL Plex Goat- $\alpha$ -Rabbit IgG-Cy3 or ECL Plex Goat- $\alpha$ MouseIgG-Cy5(Amersham) was added. After that, the membrane was washed for 3 times for 5 minutes with washing buffer. The blotting as well as all incubation and washing steps were performed at 4 °C. For chemiluminescent detection, the membrane was incubated for 10 min. in Super Signal R West Pico Chemiluminescent substrate (ThermoScientific) or for 1 - 5 min. in Super Signal™ West Fem to Maximum Sensitivity Substrate. Chemiluminescence was detected by Multi image™ Flur Chem FC2 imager (Alpha Innotech). For fluorescence detection, the membrane was

rinsed with H<sub>2</sub>O, completely air-dried in the dark and imaged with the laser scanner Typhoon FLA 9500 (GE Healthcare).

### 7.4.5 SG-tagged Protein purification

The Affinity purification of GS-tagged proteins followed by mass spectrometry analysis was performed according to Pfab et al., 2017. 15 g of transformed PSB-D cells were frozen in liquid nitrogen and ground to a fine powder using mortar and pestle. The ground cells were mixed with 20 mL prechilled extraction Buffer (25 mM HEPES-KOH pH 7.4, 100 mM NaCl, 0.05 % (v/v) IGEPALCA-630, 1 mM DTT, 2 mM MgCl<sub>2</sub>, 5 mM EGTA, 10 % (w/v) glycerol, cOmpleteTMEDTA free proteinase inhibitor tablets (Sigma-Aldrich), 1 mM PMSF dissolved in 2-propanol). The semi thawed slurry of both Falcon tubes was pooled (Total volume of 30ml) and kept on ice. Defrosted cells were sonicated 5 times for 30 secs at 30 % intensity followed by 60-seconds intervals for cooling using a UW2070 MS73 (Bandelin) Sonicator to disrupt the cells and shear the nucleic acids. MgCl<sub>2</sub> (to a final concentration of 5 mM) and 50 U/mL Benzonase were added to the mixture to degrade RNA and DNA. The sample was incubated for 30 minutes at 4 °C on a rotating wheel to allow degradation of nucleic acids. Cell debris was removed by centrifugation at 40.000 x g at 4 °C for 60 minutes. The supernatant was filtered through a 0.45 µm syringe filter. 50 µL of filtered solution was kept as an input sample. 150 µL of IgG-magnetic beads were washed four times with extraction buffer and added to the protein extract. The mixture was incubated for 1.5 h at 4 °C on a rotating wheel. Beads were pelleted by centrifugation at 2000 g and 4 °C for 5 min and subsequently transferred into new Eppendorf tube after supernatant removal. The beads were washed three times with 1 mL extraction buffer using the magnetic rack. Proteins were eluted by adding 300 µL elution buffer (0.1 M glycine-HCl, adjusted to pH 2.7) to the washed beads and incubating the mixture for 5 minutes at RT shaking at 700 rpm. After pelleting the beads with the magnetic rack, the supernatant that contains the purified proteins was transferred to new Eppendorf tube. The eluate was precipitated with ice-cold acetone and resuspended in 24 µL 1x PBS. 6 µL of 6x SDS loading buffer (150 mM Tris-HCl pH 7.0, 150 mM DTT, 5% SDS (w/v), 25% glycerol (v/v), and 0.1% bromphenol blue (w/v)) was added and samples were desaturated at 95 °C for 10 min. Obtained AP eluates were stored in -20 °C.

### 7.4.6 Trypsin in gel-digestion and purified proteins

After purification, SDS gel electrophoresis and staining with Coomassie Brilliant Blue (as described before), proteins were cut out from the gel with a scalpel. One gel lane was divided into 8 gel pieces and each piece was cut into 1mm fragments. Fragments were transferred into a 2 mL tube and washed four times while gently shaking. 1x 60 min (50 mM NH<sub>4</sub>HCO<sub>3</sub>), 1x 60 min (Mix 50 mM NH<sub>4</sub>HCO<sub>3</sub> and acetonitrile in the ratio 3:1), 1x 30 min (Mix 50 mM NH<sub>4</sub>HCO<sub>3</sub> and acetonitrile in the ratio 1:1) and 1x 10 min (200 µL acetonitrile). To reduce cysteines, 200 µL of 1 mg/ml DTT dissolved in 50 mM NH<sub>4</sub>HCO<sub>3</sub> were added to the samples and incubated for 35 min at 56 °C. For carbamidomethylation of the cysteines, the supernatant was removed and 200 µL of 5 mg/ml iodoacetamide dissolved in 50 mM NH<sub>4</sub>HCO<sub>3</sub> was added to the samples. Samples were incubated at RT in the dark for 35 min. After the repetition of the four washing steps, gel fragments were lyophilised. Gel pieces were next dried by 1 h lyophilisation and proteins were in-gel digested by the addition of trypsin (Roche), 0.04 µg/1 µL in 50 mM NH<sub>4</sub>HCO<sub>3</sub>). Dry gel fragments were incubated with 10-20 µL of trypsin mix (0.04 µg/µL trypsin in 50 mM NH<sub>4</sub>HCO<sub>3</sub>), covered with 40 µL 50 mM NH<sub>4</sub>HCO<sub>3</sub> and incubated o/n at 37 °C. The supernatant that contains the extracted peptides was

transferred to a 0.5 mL collection tube. 40  $\mu$ L 100 mM  $\text{NH}_4\text{HCO}_3$  was added to each tube and the samples were incubated for 60 min at 39 °C while gently shaking. The supernatant was transferred to the collection tube and the extraction step was repeated once with 100 mM  $\text{NH}_4\text{HCO}_3$  at 39 °C and once with a mix of 100 mM  $\text{NH}_4\text{HCO}_3$  and acetonitrile in the ratio 1:1 at 30 °C. The collection tube with the pooled peptides of all three extraction steps was lyophilized o/n.

## Mass spectrometry

Mass spectrometry was performed in the lab of Dr. Astrid Bruckmann (Department of Biochemistry I, University of Regensburg) as described in Antosz et al. 2017. Peptides obtained by trypsin digestion were resolved by reversed phase chromatography using UltiMate 3000 RSLCnano System (Thermo Scientific) and Reprosil-Pur Basic C18 nano column. A linear gradient of 4 to 40% acetonitrile in 0.1% formic acid was applied for 90 min. The HPLC system was coupled to a maXis plus UHR-QTOF system (Bruker Daltonics) via a nanoflow electrospray source (Bruker Daltonics). Data-dependent acquisition of tandem mass spectrometry (MS/MS) spectra by CID fragmentation was performed utilizing a dynamic method with a fixed cycle time of 3 s (Compass 1.7; Bruker Daltonics). Protein Scape 3.1.3 (Bruker Daltonics) in connection with Mascot 2.5.1 (Matrix Science) facilitated database searching of the NCBI database. Mascot peptide ion score cut-off was set to 25. A protein score of minimum 80 was considered as the criteria for reliable protein identification.

## 6.6 Microbial work

### 7.6.1 Cultivation of bacteria

The standard growth medium for all the bacteria strain used in this study was Luria Bertani (LB) medium (5 g/L Yeast extract, 10 g/L NaCl, 10 g/L tryptone). 1.5 % (w/v). Antibiotics were added to the following final concentrations: 100  $\mu$ g/ml ampicillin dissolved in H<sub>2</sub>O, 50  $\mu$ g/ml gentamycin dissolved in H<sub>2</sub>O, 50  $\mu$ g/ml kanamycin (*E. coli*) or 25  $\mu$ g/ml (*A. tumefaciens*) dissolved in H<sub>2</sub>O, 50  $\mu$ g/ml rifampicin dissolved in methanol, 12 $\mu$ g/ml tetracyclin dissolved in ethanol. *E. coli* and *A. tumefaciens* bacteria strains were grown at 37 °C and 28 °C, respectively. Bacteria in liquid LB media were grown under agitation at 200 rpm.

### 7.6.2 Preparation of chemically competent cells

A starter colony with a single *E. coli* or *A. tumefaciens* bacteria colony was incubated o/n at 37 °C or 28 °C and 200 rpm, in 10 mL LB medium with appropriate antibiotics. The starter culture was used to inoculate 100 mL selective LB media to an OD<sub>600</sub> of 0.1. After the cells were grown to an OD<sub>600</sub> of 0.75 they were harvested by centrifugation for 10 min at 4000 g and 4 °C. The cell pellet was re-suspended in 30 mL sterile filtered cold TBF1 buffer (100 mM RbCl, 10 mM CaCl<sub>2</sub>, 50mM MnCl<sub>2</sub>, 30mM NaOAc; adjusted to pH 5.8 with acetic acid) and incubated on ice for 90 min. The cells were harvested for 10 min at 3000 g and 4 °C and re-suspended in 4 mL of sterile filtered cold TBF2 buffer (10 mM MOPS, 10 mM RbCl, 75 mM CaCl<sub>2</sub> and 15% (w/v) glycerol). Chemically competent cells were stored in 100  $\mu$ L aliquots at -80 °C.

### 7.6.3 Transformation of chemically competent *E. coli* cells

500 ng plasmid DNA or 10  $\mu$ L ligation products were mixed with 100  $\mu$ L chemically competent *E. coli* cells. After incubation for 20 min on ice a heat shock at 42 °C was applied for 2 min. After 10 min incubation on ice the sample was mixed with 1 mL LB medium without selection and incubated for 60 min at 37 °C and 200 rpm. Transformed cells were spread on LB plates with the appropriate selection and incubated o/n at 37 °C.

### 7.6.4 Transformation of competent *A. tumefaciens* cells

1  $\mu$ g plasmid was mixed with chemically competent *A. tumefaciens* cells. Following 5 min incubation in liquid nitrogen the heat shock was applied at 37 °C for 5 min. Then the cells were kept on ice for 10 minutes. 1 mL LB medium was added and cells were incubated in a shaking incubator at 28 °C for 2 - 4 h. Cells were pelleted by centrifugation at 2000 g for 1 min and re-suspended in 25  $\mu$ L of LB medium. Suspended cells were next plated out on LB-plates containing respective antibiotics and incubated for 48 h at 28 °C.

### 7.6.5 Production of chemically competent yeast cells

Yeast AH109 cells were transferred on YPAD plates (2.2 % (w/v) micro agar, 2% (w/v) tryptone, 1% (w/v) yeast extract, 2% (w/v) glucose and 0.004% (w/v) adenine hemisulfate) and incubated at for 3-4 days at 30 °C. As a starter culture, 3 mL YPAD (2% (w/v) tryptone, 1% (w/v) yeast extract, 2% (w/v) glucose and 0.004% (w/v) adenine hemisulfate) medium were inoculated with a single colony of AH109 yeast cells and incubated o/n under agitation at 200 rpm at 30 °C to produce chemical competent yeast cells. 50 mL YPAD medium was mixed with the starter culture to the OD600 of 0.1. Cells were subsequently grown in a shaking incubator at 30 °C and 200 rpm until an OD600 of 0.5 - 1. The cells were pelleted by centrifugation for 5 min. at 500 g and washed. First with 25 mL sterile H<sub>2</sub>O then with 5 mL sterile filtered SORB buffer (100 mM LiOAc, 10 mM Tris, 1 mM EDTA, 1 M sorbitol; adjust to pH 8.0) and last with 500  $\mu$ L SORB buffer. The washed pellet was re-suspended in 360  $\mu$ L SORB buffer and mixed with 40  $\mu$ L ice-cold denatured single stranded 10 mg/mL salmon sperm DNA. For denaturing, the salmon sperm DNA was heated at 90 °C for 5 min. and immediately kept on ice. 50  $\mu$ L aliquots of the competent yeast cells were stored at -80 °C.

### 7.6.6 Transformation of yeast competent cells

An aliquot of competent AH109 yeast cells was thawed on ice and gently mixed with 500 ng of bait (pGBKT7-derivate) and prey (pGADT7-derivate) plasmid DNA and 300  $\mu$ L sterile filtered PEG solution (100mM LiOAc, 10mM Tris, 1mM EDTA, 40% PEG3350; adjusted to pH 8). The solution was incubated for 30 min at RT. 40  $\mu$ L DMSO was added to the sample. Heat shock was subsequently applied at 42°C for 15 min. Cells were pelleted at 500 g for 2 min and re-suspended in 200  $\mu$ L H<sub>2</sub>O. The solution was plated out on selective plates and grown for 3-4 days at 30°C until colonies were visible. Selection was done on double dropout (DDO) (2% w/v Glucose, 0.67% w/v

Yeast nitrogen base w/o amino acids, 2.2% w/v agar, 0.064% w/v-Leu/Trp DO supplement; adjusted to pH 5.8) or quadruple dropout (QDO) (2% w/v Glucose, 0.67% w/v Yeast nitrogen base w/o amino acids, 2.2% w/v agar, 0.064% w/v Ade/-His/-Leu/-Trp/ DO supplement (Clonetechn: #630428) ; adjusted to pH 5.8) medium.

### 7.6.7 Yeast-2-Hybrid assay

A single positive-selected yeast colony was picked from DDO plates after the co-transformation of bait and prey plasmids and it was resuspended in 200  $\mu$ L H<sub>2</sub>O. OD<sub>600</sub> was determined with one half of the solution. Remaining yeast suspension was adjusted to an OD<sub>600</sub> of 1.0 and subjected to 1:10 dilution series (100, 10<sup>-1</sup>, 10<sup>-2</sup>, 10<sup>-3</sup>). Respective dilutions were placed on DDO and QDO plates using a frogger. Plates were incubated for 3 - 4 days at 30 °C until colonies were visible.

## 7.7 Plant Work

### 7.7.1 Cultivation of Arabidopsis plants

Soil (80% (v/v) Einheitserde Typ ED 73, 10% (v/v) supplemented with sand and 10% Isoself R from Knauf Perlite was used to sow out *Arabidopsis* seeds. The prepared mixture was soaked with water containing 0.03% (v/v) confidor WG70 (Bayer) and 3 g/L fertiliser Osmocote Start (The Scotts Company) prior to sowing out. Seeds in pots were stratified in the dark for 48 hours at 4 °C. Pots were next transferred into the growing chamber and plants were grown under long-day (LD) conditions (16 h light and 8 h dark) at 22 °C and light intensity of 120  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup>. Pots were regularly shuffled on the shelves to ensure comparable conditions. Arabidopsis seeds were sterilized using chloric gas (40 mL 12.5% hypochloric acid (w/v) and 2 mL 37% HCl (v/v)) for plants grown under sterile conditions. Arabidopsis seeds were sown on solid 0.5x MS plates (2.15 g/L Murashige and Skoog media, 1% (w/v) sucrose, 0.7% or 1% (w/v) phyto agar adjusted to pH 5.8). After stratification (2-3 days at 4 °C in the dark), plants were grown in a plant incubator (Percival Scientific) under LD conditions at 22 °C and 100  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup>.

### 7.7.2 Transformation of Arabidopsis plants by floral dipping

*Arabidopsis* plants were stably transformed using the Floral Dip method described in Clough et al. 1998. Briefly, approximately 20 Arabidopsis plants with 10-15 cm high inflorescences were used for each transformation. *A. tumefaciens* GV3101 + pSoup carrying cells transformed by heat shock with pGreen-derived plasmids or pCambia were used for transformation. The insertion of the plasmid was tested by colony PCR. 5 mL LB medium with selection were initially inoculated as starter cultures and grown overnight. 500 mL selective LB were inoculated with 500  $\mu$ L of the starter culture and incubated o/n at 28 °C and 200 rpm. The cells were harvested by centrifugation for 15 min at 5000 g and were re-suspended in 500 mL infiltration medium (5% (w/v) sucrose, 10 mM MgCl<sub>2</sub>, 0.02% (v/v) Silvet L77 and 0.01 mM acetosyringone dissolved in ethanol). The inflorescences of the plants were dipped into the infiltration medium for 1 min and were left o/n covered with plastic foil. Plants were grown to maturity in the plant growth chamber and plants of the next generation were selected for transformation with the transgene.

### 7.7.3 Crossing of *Arabidopsis* plants

To obtain double mutants the open flowers and meristems of an inflorescence from *Arabidopsis* plants were removed with a tweezer. The buds with the right size were emasculated by gently removing all sepals, petals and stamens. The pollen from the genotype B was transferred on the stigma of genotype A and the inflorescence was labelled. After 14-21 days in the plant growth chamber, the silique with the hybrid seeds was harvested.

### 7.7.4 Soil-based phenotyping of *Arabidopsis* plants

For the phenotypic analysis, the *Arabidopsis* plants were grown under long day condition in a growth chamber. To avoid position effects they were rotated every 2 days in the shelf. Bolting time (the emergence of the flower bud), the rosette diameter (at bolting), the number of secondary inflorescence (at 42 DAS) and plant height (at 42 DAS) were monitored. All pictures were taken with EOS 600D equipped with a Macro lens EF-S 60 mm 1:2.8 USM (Canon) or ETS 18-55 mm objective (Canon).

### 7.7.5 *Arabidopsis* plants grown on plates for alternative splicing analysis

Before sowing out, seeds were surface sterilized in the exsiccator with chloric gas (40 mL 12.5% hypochloric acid (w/v) and 2 mL 37% HCl (v/v)). To keep plant growth under sterile condition, the sterile seeds were immediately sown out on solid 1x MS plates (4.3 g/L Murashige and Skoog media including vitamins, 1% sucrose, 0.8% phyto agar (w/v); adjusted to pH 5.8). Plates were secured with Micropore surgical tape and kept for 2 days in dark conditions at 4 °C for seeds stratification. Plates were next transferred into the plant incubator (Percival Scientific) under LD conditions at 22 °C with light intensity 120  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Plates were regularly shuffled on the shelves to ensure comparable conditions between the different genotypes. For the selection of transgenic plants MS medium was supplemented with 30  $\mu\text{g/mL}$  Hygromycin B.

### 7.7.6 Cultivation of *Arabidopsis* PSB-D cells

The *Arabidopsis* landsberg erecta PSB-D suspension cells (Arabidopsis Biological Resource Center) were cultured and transformed under sterile conditions according to Van Leene et al., 2011. The suspension cells culture was grown in darkness at 25°C in MSMO medium (0.443 % Murashige and Skoog Salt mixture (US Biological), 3 % sucrose, 0.5 mg/L NAA dissolved in 100 mM NaOH, 100 mg/L myo-inositol, 0.05 mg/L kinetin dissolved in DMSO, 0.4 mg/L thiamine, adjusted to pH 5.7 with 1 M KOH) while shaking at 130 rpm. Once a week, the cells were diluted by transferring 7 mL cells into a 100 mL Erlenmeyer flask containing 43 mL MSMO medium. For selection, MSMO medium was supplemented with kanamycin or hygromycin B to a final concentration of 50  $\mu\text{g/mL}$  or 20  $\mu\text{g/mL}$ , respectively.

### 7.7.7 Transformation of *Arabidopsis* PSB-D cells

The *Arabidopsis* suspension cells were transformed by *A. tumefaciens* mediated cell transformation (8.1.4). The vectors were introduced into the bacteria by heat stress. 2 mL LB medium with selection was initially inoculated and culture was grown overnight. Next 20 mL LB medium with selection was inoculated with an overnight culture. After 24 h growth *Agrobacterium* cells were pelleted by centrifugation at 3000 g for 15 min and re-suspended in 40 mL sterile MSMO medium. The washing step was repeated once again and the OD600 of the cell suspension was adjusted to 1.0. Then 3 mL of 3-days old *Arabidopsis* suspension cells (OD600: 1.2 - 1.3) were co-cultivated with 200  $\mu$ L of the *A. tumefaciens*/MSMO solution and 6  $\mu$ L 100 mM acetosyringone in one well of a 6-well plate. The Plate was secured with Micropore surgical tape and transferred into the shaking incubator at 130 rpm for 3 days in a and 25 °C in the dark. For selection in liquid medium, the 3 mL transformed cells were transferred into a 25 mL Erlenmeyer flask containing 8 mL MSMO medium supplemented with 50  $\mu$ L/mL kanamycin or 20  $\mu$ g/mL hygromycin B (for plasmid selection), 500  $\mu$ L/mL vancomycin and 500  $\mu$ L/mL carbenicillin (to kill *A. tumefaciens* cells). *Arabidopsis* PSD-B culture was subsequently grown in the shaking incubator for 7 days. Next, sedimented cells were transferred into a 100 mL Erlenmeyer flask containing 35 mL MSMO with kanamycin or hygromycin, vancomycin and carbenicillin. In the following suspension cultures were diluted once a week.

For selection on solid medium, the ~3 mL of transformed cells were washed with 40 mL MSMO medium in a 50 mL Falcon tube. After centrifugation (5 min., 500 g), 2-3 mL of the harvested cells were spread on MSMO plates supplemented with 50  $\mu$ L/mL kanamycin, 500  $\mu$ L/mL vancomycin and 500  $\mu$ L/mL carbenicillin. The MSMO plates were incubated for 2 - 3 week at 25 °C in the dark. Using a scapel, the grown callus was scraped off the plate, chopped and transferred into a 100 mL Erlenmeyer flask containing 30 mL MSMO supplemented with 50  $\mu$ L/mL kanamycin, 500  $\mu$ L/mL vancomycin and 500  $\mu$ L/mL carbenicillin. After one week of incubation at 25 °C in the shaking incubator, the transformed cells were diluted regularly once a week.

The transformed *Arabidopsis* suspension cells were gradually upscaled by transferring into increasing volume of MSMO medium every week. After 4-8 weeks, cell culture from each flask was transferred into 2 L Erlenmeyer flask containing 800 mL MSMO medium. Following 3 days of cultivation in a shaking incubator cells were harvested by filtering the cell suspension through the double layer of Miracloth. Cells were subsequently harvested in 15 g aliquots and stored at -80 °C.

Cryo-stocks were generated for long term storage of *Arabidopsis* suspension cells. 10 mL of 7-days old cell culture were transferred into 100 mL Erlenmeyer flasks containing 40 mL pre-freeze medium (0.443 % Murashige and Skoog Salt mixture (US Biological), 3 % sucrose, 2.7% mannitol, 0.5mg/LNAA dissolved in 100m MNaOH, 100mg/L myo-inositol, 0.05mg/L kinetin dissolved in DMSO, 0.4 mg/L thiamine; adjusted to pH 5.7 with 1 M KOH and autoclave) and was incubated for two days at 25 °C and 200 rpm. The pre-freeze culture was harvested by 1min centrifugation at 130g in a 50 mL falcon tube. The supernatant was removed and the cell pellet (~ 2-3 mL) was mixed with 6 mL cryoprotective media (0.443 % Murashige and Skoog Salt mixture (US Biological), 8.56 % sucrose, 1.15 % glycerol, 0.25 % proline, 0.5 mg/L NAA dissolved in 100 mM NaOH, 100 mg/L myo-inositol, 0.05 mg/L kinetin dissolved in DMSO, 0.4 mg/L thiamine; adjusted to pH 7.0 with 1 M KOH and filter sterilize). The mixture was incubated for one hour on ice. The cells in cryoprotective media were aliquoted (2 mL) in ice cold 2 mL Nalgene tubes. The tubes were placed for 2-3 hours at -80 °C, then frozen in liquid nitrogen and stored in a cryogenic container filled with liquid nitrogen.

### 7.7.8 Infiltration of *Nicotiana benthamiana* leaves

10 mL of selective LB medium were incubated by co-cultivation with *A. tumefaciens* harbouring a plant expression vector over night at 28°C and 200 rpm. After centrifugation for 10 min at 4000 g the bacteria pellet was resuspended in 10 mL infiltration medium (10 mM MES-KOH pH 5.7, 10 mM MgCl<sub>2</sub>, 0.1 mM acetosyringone dissolved in ethanol). *A. tumefaciens* cells in infiltration medium were infiltrated in the abaxial side of leaves from 2-4 weeks old *N. benthamiana* plants with a syringe. After 2-3 days in the green house, infiltrated leaves were mounted on objective slides and analysed by confocal microscopy, CLSM.

## 7.8 Microscopy

### 7.8.1 Confocal Laser Scanning Microscopy (CLSM)

Confocal laser scanning microscopy (CLSM) was performed using a Leica SP8 microscope, equipped with a 10X NA 0,3, 40X Oil NA 1,3 or 63X Glycerol NA 1.3 objective. DAPI was excited with a 405 nm laser, GFP/Alexa Fluor® 488 were excited using an Argon laser at 488 nm and tagRFP/mCherry/PI were excited using a DPSS laser at 561 nm. The emission of GFP/ Alexa Fluor® 488 and tagRFP/mCherry were detected with Hybrid detectors at 500-550 nm or 570-620 nm, respectively. The emissions of DAPI or PI were detected with PMTs at 410/495 nm or 570-620 nm, respectively. Pictures in this study were generated using ImageJ software version 1.49.

### 7.8.2 Förster resonance energy transfer (FRET)

Fragments of *N. benthamiana* leaves (3 - 4 weeks old) were mounted in H<sub>2</sub>O on objective slides with cover slips with the abaxial side facing up and FRET efficiencies were measured and calculated as described in Weidtkamp-peters and Stahl, 2017. For FRET acceptor photobleaching (FRET-APB), eGFP and mCherry fusion proteins acting as donor/acceptor pairs were expressed in *N. benthamiana* leaves (9.1.8). Images were acquired using a SP8 (Leica, Wetzlar, Germany) confocal laser scanning microscope (CLSM) equipped with a 63X Glycerol NA 1.3 objective. GFP was excited using an Argon laser at 488 nm and mCherry was excited and bleached using a DPSS laser at 561 nm. For bleaching, a circular ROI area of 10 µm was defined and it was bleached at 100% laser power for 80 iterations. 15 pre-bleach and 15 post-bleach images were analysed by ImageJ software version 1.49m. The mean FRET efficiencies were calculated by the following formula:  $([I_{POST} - I_{PRE}] / I_{PRE}) \times 100$  with  $I_{POST}$  = mean fluorescence intensity of 15 postbleached frames and  $I_{PRE}$  = mean fluorescence intensity of 15 prebleached frames.



## 8. Supplements

**Supplementary table 1a. Differentially alternatively spliced genes in *tef* mutants.** List of 167 significant expressed alternatively spliced genes in the mutant genotype *spt4R3*, *tflls*, *iws1*, *tflls/iws1*, according to RNAseq. In green are respectively marked the statistically significant genes with  $pval < 0.05$  in the mutant analysed.

ATG	Description	Spt4R3	tflls	iws1	tflls/iws1	AS event
AT5G17230	phytoene synthase	6,6E-07	1,2E-01	1,3E-03	2,0E-02	alt_5prime
AT2G38870	PR-6 proteinase inhibitor	3,7E-05	4,4E-01	8,9E-02	1,0E+00	alt_5prime
AT3G22210	transmembrane protein	1,3E-04	4,0E-02	5,9E-01	4,3E-01	alt_5prime
AT3G61860	RS31	3,5E-04	5,1E-02	8,5E-06	1,9E-05	intron_retention
AT1G69840	ATHIR1,	4,6E-04	1,0E+00	2,0E-01	5,9E-01	intron_retention
AT3G26450	cyclase/dehydrase	7,4E-04	5,0E-01	2,2E-03	2,6E-02	alt_3prime
AT5G66190	LFNR1, ATLFNR1, FNR1	1,3E-03	2,8E-03	8,5E-04	7,6E-04	intron_retention
AT2G21970	SEP2	3,3E-03	5,5E-02	4,2E-03	2,5E-01	intron_retention
AT4G31580	SRZ-22	3,9E-03	3,0E-02	3,3E-02	2,1E-02	intron_retention
AT5G58350	ZIK2, WNK4	5,2E-03	1,6E-01	3,5E-02	1,4E-01	alt_5prime
AT1G70770	transmembrane protein	5,5E-03	2,0E-01	3,0E-02	4,3E-03	intron_retention
AT2G04039	NDHV	5,9E-03	6,3E-02	3,0E-01	2,0E-01	alt_5prime
AT4G26640	AtWRKY20, WRKY20	6,1E-03	3,1E-01	3,0E-01	2,9E-01	alt_5prime
AT1G76990	ACR3	1,0E-02	3,0E-01	8,6E-03	1,2E-02	alt_5prime
AT1G79550	PGK	1,0E-02	1,8E-02	1,0E-03	9,5E-02	intron_retention
AT1G07820	Histone superfamily protein	1,4E-02	2,5E-02	1,3E-02	5,7E-02	intron_retention
AT2G44180	Encodes a MAP2 like	1,5E-02	1,8E-01	1,5E-01	1,8E-01	alt_3prime
AT3G52590	UBQ1	1,5E-02	1,8E-02	3,5E-02	9,2E-02	alt_3prime
AT5G05360	hypothetical protein;	1,5E-02	5,4E-02	5,7E-03	8,3E-02	alt_5prime
AT2G14910	MAR-binding filament	1,6E-02	7,8E-02	4,3E-02	1,7E-01	intron_retention
AT1G77180	SKIP	1,6E-02	1,2E-01	3,4E-02	5,6E-02	intron_retention
AT2G39550	ATGGT-IB, PGGT-I, GGB	1,8E-02	1,2E-01	2,3E-01	2,4E-01	alt_3prime
AT1G47410	hypothetical protein	2,2E-02	5,8E-01	7,8E-01	7,7E-01	intron_retention
AT3G20870	ZTP29	2,2E-02	5,2E-01	6,2E-02	1,2E-01	alt_3prime
AT1G52230	PSI-H, PSAH-2, PSAH2	2,3E-02	2,6E-02	9,1E-02	1,7E-01	alt_5prime
AT3G43520	Transmembrane proteins 14C	2,3E-02	3,8E-02	4,3E-03	2,5E-02	intron_retention
AT2G42950	Mg transporter CorA-like	2,3E-02	3,2E-01	4,8E-01	6,7E-01	alt_3prime
AT4G00480	myc1, ATMYC1	2,5E-02	2,8E-01	3,1E-02	3,0E-01	intron_retention
AT3G03100	NADH	2,8E-02	2,2E-01	2,3E-01	3,8E-01	alt_3prime
AT4G02790	ATRBGA	2,8E-02	4,0E-01	4,3E-01	4,8E-01	alt_5prime
AT2G31750	auxin glycosyltransferase	2,9E-02	1,3E-02	2,3E-01	2,0E-02	alt_5prime
AT3G13060	ECT5	3,0E-02	3,7E-01	2,1E-02	1,7E-01	alt_5prime
AT5G35170	adenylate kinase	3,0E-02	8,9E-03	5,8E-02	2,5E-01	alt_5prime
AT4G21860	MSRB2	3,1E-02	2,6E-02	2,0E-02	2,1E-01	intron_retention
AT5G13480	FY	3,2E-02	1,6E-01	4,2E-01	2,3E-01	alt_3prime
AT3G05727	DEG28	3,5E-02	3,6E-02	2,8E-02	4,5E-01	alt_3prime
AT1G06700	Protein kinase	3,6E-02	1,2E-02	1,5E-01	3,3E-01	intron_retention
AT3G63140	CSP41A	3,7E-02	2,1E-02	1,0E-01	1,1E-01	alt_5prime
AT3G25840	PRP4 KINASE A, PRP4KA	3,7E-02	1,0E-01	5,9E-02	1,1E-01	alt_3prime
AT1G25054	ATLPXC3,	3,7E-02	3,6E-01	1,6E-03	4,1E-02	alt_5prime
AT1G12140	FMO GS-OX5	3,8E-02	4,1E-01	2,3E-01	2,2E-01	alt_3prime
AT1G12140	FMO GS-OX5	3,8E-02	4,1E-01	2,3E-01	2,2E-01	alt_3prime
AT1G22060	sporulation-specific protein	3,9E-02	1,7E-01	1,7E-01	1,2E-01	intron_retention
ATCG00920	RIBOSOMAL RNA16S,	3,9E-02	3,1E-02	7,4E-02	3,7E-01	alt_5prime

**Supplementary table 1b. Differentially alternatively spliced genes in *tef* mutants.** List of 167 significant expressed alternatively spliced genes in the mutant genotype *spt4R3*, *tflls*, *iws1*, *tflls/iws1*, according to RNAseq. In green are respectively marked the statistically significant genes with *pval* < 0.05 in the mutant analysed.

ATG	Description	Spt4R3	tflls	iws1	tflls/iws1	AS event
AT3G14220	GDSL-motif esterase	3,9E-02	3,3E-01	3,1E-01	3,8E-01	alt_3prime
AT1G52510	alpha/beta-Hydrolases	4,0E-02	1,2E-01	1,3E-02	1,7E-01	alt_5prime
AT3G02750	Protein phosphatase 2C	4,1E-02	1,5E-02	2,6E-01	1,9E-03	intron_retention
AT3G24570	Peroxisomal membrane	4,1E-02	3,1E-01	5,0E-02	1,5E-01	intron_retention
AT1G01490	Heavy metal transport	4,2E-02	5,7E-01	5,6E-01	2,3E-01	intron_retention
AT5G64170	LNK1	4,3E-02	2,4E-01	3,1E-01	1,7E-01	alt_3prime
AT5G18420	CCR4-NOT	4,3E-02	1,3E-01	1,6E-01	5,1E-02	exon_skip
AT5G38420	RBCS2B	4,3E-02	2,4E-02	3,1E-02	8,2E-02	intron_retention
AT2G37620	ACT1, AAc1	4,4E-02	3,2E-01	2,5E-01	1,8E-01	alt_3prime
AT4G24060	Dof-type zinc finger D	4,5E-02	6,4E-02	2,1E-01	9,6E-02	intron_retention
AT3G58610	ketol-acid reductoisomerase	4,6E-02	6,3E-02	1,9E-02	1,3E-01	alt_5prime
AT3G16520	UDP-GLUCOSYL TRANS	4,6E-02	1,0E-01	1,2E-01	7,4E-01	alt_5prime
AT1G45474	Lhca5	1,4E-01	4,9E-02	4,1E-02	2,3E-01	alt_3prime
AT3G54350	emb1967	3,0E-01	5,0E-02	2,0E-01	2,7E-01	alt_3prime
AT3G58510	RH11, RNA HELICASE 11	9,7E-02	1,1E-01	1,6E-05	1,8E-04	intron_retention
AT3G49430	SR34a, SRp34a, At-SR34a	2,3E-01	4,1E-01	3,9E-04	5,2E-02	intron_retention
AT5G25280	SR protein	9,6E-01	8,7E-01	6,1E-04	3,7E-02	alt_5prime
AT1G19400	E-4PDHAS	2,4E-01	1,5E-01	2,1E-03	2,2E-01	alt_5prime
AT3G28270	AFL1	6,5E-01	1,5E-01	2,9E-03	2,7E-05	intron_retention
AT3G29575	AFP3	5,2E-01	9,4E-02	3,7E-03	7,7E-04	intron_retention
AT3G04500	RNA-binding	2,2E-01	2,4E-01	5,1E-03	5,2E-02	intron_retention
AT2G28470	putative beta-galactosidase	3,3E-01	4,1E-01	1,0E-02	1,3E-01	exon_skip
AT2G22090	UBA1A	6,8E-02	5,6E-01	1,0E-02	6,8E-03	intron_retention
AT2G45850	AHL9	1,8E-01	2,0E-01	1,1E-02	3,3E-01	intron_retention
AT5G02940	ion channel POLLUX	1,1E-01	9,4E-02	1,3E-02	3,8E-01	intron_retention
AT3G55390	CASPL4C1	2,9E-01	1,7E-01	1,4E-02	3,7E-01	alt_5prime
AT2G45960	PLASMA MEMBRANE	2,4E-01	2,2E-01	1,4E-02	1,9E-02	alt_5prime
AT5G18230	NOT3	4,2E-01	9,7E-02	1,5E-02	1,3E-01	exon_skip
AT5G07370	INOSITOL POLYPHOSPHATE	1,8E-01	1,8E-01	1,6E-02	1,6E-01	intron_retention
AT1G68010	HPR, ATHPR1	1,2E-01	6,9E-02	1,6E-02	1,8E-01	alt_3prime
AT3G26290	CYP71B26	1,7E-01	4,4E-01	1,7E-02	2,2E-01	intron_retention
AT1G76670	URGT1	1,0E+00	3,4E-01	1,8E-02	2,5E-03	intron_retention
AT5G19480	no symbol available	2,7E-01	6,5E-01	1,8E-02	9,6E-01	alt_5prime
AT5G58140	PHOT2, AtPHOT2, NPL1	3,4E-01	3,6E-01	1,8E-02	3,4E-02	alt_3prime
AT3G18050	GPI-anchored protein	2,0E-01	5,5E-01	1,9E-02	7,8E-01	intron_retention
AT3G57530	ATCPK32	1,7E-01	3,6E-01	1,9E-02	1,4E-02	alt_3prime
AT3G26680	ATSNM1, SNM1	1,9E-01	1,6E-01	2,0E-02	9,9E-02	alt_5prime
AT1G31190	IMPL1	9,0E-02	1,1E-01	2,0E-02	5,7E-02	alt_3prime
AT5G35160	TMN11	5,9E-01	7,2E-01	2,0E-02	5,0E-01	intron_retention
AT5G27680	RECQSIM	1,8E-01	1,9E-01	2,1E-02	1,8E-01	alt_5prime
ATCG00730	PHOT. ELECTRON TRAN.	2,0E-01	9,3E-02	2,2E-02	3,1E-03	intron_retention
AT3G62290	ARFA1E, ATARFA1E	4,1E-01	2,7E-01	2,2E-02	1,6E-01	alt_5prime

**Supplementary table 1c. Differentially alternatively spliced genes in *tef* mutants.** List of 167 significant expressed alternatively spliced genes in the mutant genotype *spt4R3*, *tflls*, *iws1*, *tflls/iws1*, according to RNAseq. In green are respectively marked the statistically significant genes with *pval* < 0.05 in the mutant analysed.

ATG	Description	Spt4R3	tflls	iws1	tflls/iws1	AS event
AT3G62290	ARFA1E, ATARFA1E	4,1E-01	2,7E-01	2,2E-02	1,6E-01	alt_5prime
AT3G44310	NIT1, ATNIT1, NITI	1,1E-01	1,4E-01	2,4E-02	8,7E-02	alt_5prime
AT3G01500	CA1	3,2E-01	7,7E-02	2,4E-02	1,2E-01	alt_3prime
AT4G39710	PnsL4, FKBP16-2	5,4E-02	1,8E-01	2,4E-02	8,8E-02	exon_skip
AT2G14260	proline iminopeptidase	6,3E-01	6,0E-01	2,5E-02	6,0E-01	exon_skip
AT2G25970	no symbol available	2,8E-01	2,2E-01	2,5E-02	7,1E-03	alt_5prime
AT4G34135	UGT73B2	2,9E-01	9,0E-01	2,5E-02	1,2E-01	intron_retention
AT1G36160	AT-ACC1	2,5E-01	5,6E-01	2,6E-02	2,8E-01	alt_5prime
AT1G01290	CNX3	6,7E-02	7,0E-01	2,7E-02	5,7E-03	intron_retention
AT4G37270	ATHMA1, HMA1	6,2E-01	3,0E-01	2,8E-02	2,3E-01	alt_5prime
AT5G11500	coiled-coil protein	5,4E-02	1,2E-01	2,9E-02	4,7E-02	alt_5prime
AT5G09880	Splicing factor, CC1	3,8E-01	4,5E-01	2,9E-02	6,6E-02	alt_5prime
AT1G02840	SR34	1,7E-01	3,7E-01	2,9E-02	7,0E-03	alt_5prime
AT5G42080	DRP1A,	8,2E-02	2,6E-01	3,0E-02	3,4E-01	alt_3prime
AT5G18100	CSD3	2,3E-01	2,3E-01	3,1E-02	6,0E-02	intron_retention
AT4G36690	ATU2AF65A	9,7E-02	3,5E-01	3,2E-02	9,7E-02	alt_3prime
AT1G72970	EDA17, HTH	5,3E-02	1,7E-01	3,3E-02	5,6E-01	alt_3prime
AT1G76180	ERD14	1,5E-01	8,6E-02	3,4E-02	2,9E-03	intron_retention
AT5G27840	TOPP8	1,1E-01	5,2E-02	3,4E-02	2,0E-02	intron_retention
AT3G56240	CCH	9,6E-02	9,3E-02	3,5E-02	2,4E-02	alt_5prime
AT4G35450	ANKYRIN REPEAT	1,3E-01	6,2E-02	3,5E-02	6,6E-02	alt_5prime
AT3G10230	AtLCY, LYC	2,7E-01	5,0E-01	3,5E-02	2,3E-01	intron_retention
AT4G21150	RPN2, HAP6	3,0E-01	3,8E-01	3,6E-02	2,5E-01	alt_5prime
AT1G61520	LHCA3	1,0E-01	1,1E-01	3,7E-02	3,1E-01	alt_3prime
AT2G42130	PAP	3,2E-01	4,4E-01	3,7E-02	1,3E-01	alt_3prime
AT4G10340	LHCB5	1,2E-01	3,4E-01	4,0E-02	8,8E-03	alt_5prime
AT4G16360	KINβ2	5,2E-01	6,5E-01	4,1E-02	5,6E-01	alt_3prime
AT1G08550	AVDE1, NPQ1	4,7E-01	2,6E-01	4,2E-02	4,6E-01	alt_5prime
AT5G60390	EF1alpha	9,5E-02	8,3E-02	4,3E-02	2,1E-01	intron_retention
AT5G04220	NTMC2TYPE1	2,0E-01	7,0E-01	4,3E-02	3,5E-01	alt_5prime
AT4G34700	ATCIB22	3,5E-01	2,9E-01	4,5E-02	3,1E-01	alt_3prime
AT3G59950	Peptidase family C54	1,7E-01	3,2E-01	4,5E-02	6,4E-02	alt_3prime
AT5G42220	Ubiquitin	6,1E-01	8,5E-01	4,6E-02	2,6E-01	alt_5prime
AT5G48760	Ribosomal protein L13	3,7E-01	3,2E-01	4,7E-02	7,1E-01	alt_5prime
AT3G13080	MRP3, ATMRP3, ABCC3	3,8E-01	2,7E-01	4,8E-02	4,1E-01	alt_3prime
AT3G29390	RIK	2,0E-01	5,2E-01	4,8E-02	2,8E-01	alt_5prime
AT3G02730	TRXF1, ATF1	8,2E-02	6,0E-02	4,8E-02	2,6E-01	intron_retention
AT1G80420	XRCC1, ATXRCC1	2,2E-01	5,5E-01	4,8E-02	5,0E-01	alt_3prime
AT5G13110	GLUCOSE-6-PHOS. DEHY. 2	5,7E-01	5,1E-01	4,9E-02	3,9E-01	alt_3prime
AT1G68490	transl. subunit seca	6,1E-01	7,3E-01	2,6E-01	6,4E-04	intron_retention
AT1G05090	phoprotein-like protein	3,9E-01	2,5E-01	9,6E-01	4,7E-03	intron_retention
AT4G39680	SAP domain	1,2E-01	2,2E-01	2,5E-01	5,2E-03	alt_5prime
AT2G16365	PCH1	3,3E-01	8,2E-01	7,7E-01	5,6E-03	intron_retention
AT3G46780	PTAC16	4,2E-01	2,6E-01	4,2E-01	7,1E-03	alt_3prime
AT4G03260	MASP1	7,9E-01	4,1E-01	2,5E-01	9,7E-03	alt_5prime

**Supplementary table 1d. Differentially alternatively spliced genes in *tef* mutants.** List of 167 significant expressed alternatively spliced genes in the mutant genotype *spt4R3*, *tflIs*, *iws1*, *tflIs/iws1*, according to RNAseq. In green are respectively marked the statistically significant genes with  $pval < 0.05$  in the mutant analysed.

ATG	Description	Spt4R3	tfls	iws1	tfls/iws1	AS event
AT4G03260	MASP1	7,9E-01	4,1E-01	2,5E-01	9,7E-03	alt_5prime
AT2G20890	THF1, PSB29	2,4E-01	1,8E-01	6,8E-02	1,5E-02	alt_3prime
AT3G51950	single CCCH domain	4,2E-01	2,8E-01	4,9E-01	1,7E-02	alt_3prime
AT2G46800	ATMTP1, MTP1	1,4E-01	6,7E-01	3,0E-01	2,1E-02	alt_5prime
AT5G20380	PHT4;5	3,0E-01	5,6E-01	4,2E-01	2,2E-02	alt_5prime
AT5G63370	CDKG1	4,8E-01	2,8E-01	3,4E-01	2,3E-02	alt_5prime
AT1G31910	GHMP kinase	1,3E-01	2,9E-01	4,0E-01	2,3E-02	intron_retention
AT1G56190	CPGK2	3,6E-01	3,7E-01	2,9E-01	2,5E-02	alt_5prime
AT1G11280	S-locus lectin	8,8E-01	4,5E-01	6,5E-01	2,8E-02	alt_3prime
AT4G08920	OOP2, BLU1, ATCRY1	1,9E-01	5,7E-01	3,9E-01	2,9E-02	intron_retention
AT5G53300	UBC10, AtUBC10	2,9E-01	5,0E-01	3,6E-01	3,2E-02	alt_3prime
AT5G59613	ATP synthase	2,2E-01	2,1E-01	1,0E-01	3,3E-02	exon_skip
AT5G04140	GLUS, FD-GOGAT, GLS1, GLU1	8,3E-02	9,1E-02	7,3E-02	3,4E-02	alt_5prime
AT1G62640	KAS III	2,4E-01	3,9E-01	1,4E-01	3,5E-02	intron_retention
AT2G43350	GLUTATHIONE PEROXIDASE 3	2,0E-01	2,4E-01	1,1E-01	3,6E-02	alt_3prime
AT5G59430	TRP1, ATTRP1	2,3E-01	2,1E-01	1,0E-01	3,7E-02	alt_5prime
AT3G01810	EEIG1/EHBP1	6,8E-01	6,7E-01	7,5E-02	4,0E-02	alt_5prime
AT5G53620	RNAPII degr. factor	1,0E-01	1,6E-01	1,2E-01	4,3E-02	alt_5prime
AT3G16190	Isochorismatase	4,5E-01	3,3E-01	7,2E-01	4,3E-02	alt_5prime
AT4G27090	RPL14B	3,3E-01	4,6E-01	1,6E-01	4,4E-02	alt_3prime
AT2G27530	PGY1	5,2E-01	6,6E-01	3,9E-01	4,4E-02	alt_3prime
AT3G26950	transmembrane protein	3,7E-01	3,3E-01	1,7E-01	4,5E-02	alt_3prime
AT5G06060	NAD(P)-binding Rossmann	8,1E-01	7,8E-01	8,0E-01	4,7E-02	alt_3prime
AT5G49970	ATPPOX, PPOX, PDX3	6,9E-01	6,7E-01	1,7E-01	4,7E-02	alt_3prime
AT2G27385	Pollen Ole e 1	6,3E-01	7,9E-01	3,3E-01	4,8E-02	alt_3prime

**Supplementary table 2. Gene ontology enrichment analysis of genes differentially alternatively spliced in *spt4R3* involved in metabolic process.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function. All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	metabolic process	p-value
GO:0006007	glucose catabolic process	0.000171
GO:0046365	monosaccharide catabolic process	0.000176
GO:0019320	hexose catabolic process	0.000176
GO:0044724	single-organism carbohydrate catabolic process	0.000202
GO:0044723	single-organism carbohydrate metabolic process	0.000267
GO:0006807	nitrogen compound metabolic process	0.0004
GO:0006006	glucose metabolic process	0.000413
GO:0044237	cellular metabolic process	0.000452
GO:0019318	hexose metabolic process	0.000657
GO:0044249	cellular biosynthetic process	0.00074
GO:0044699	single-organism process	0.01
GO:0044710	single-organism metabolic process	0.01
GO:0044712	single-organism catabolic process	0.01
GO:0044711	single-organism biosynthetic process	0.01
GO:0072524	pyridine-containing compound metabolic process	0.01
GO:0019362	pyridine nucleotide metabolic process	0.01
GO:0044238	primary metabolic process	0.01
GO:0006793	phosphorus metabolic process	0.01
GO:0006796	phosphate-containing compound metabolic process	0.01
GO:0006098	pentose-phosphate shunt	0.01
GO:0006733	oxidoreduction coenzyme metabolic process	0.01
GO:0019637	organophosphate metabolic process	0.01
GO:1901564	organonitrogen compound metabolic process	0.01
GO:0071704	organic substance metabolic process	0.01
GO:1901575	organic substance catabolic process	0.01
GO:1901576	organic substance biosynthetic process	0.01
GO:1901360	organic cyclic compound metabolic process	0.01
GO:0006139	nucleobase-containing compound metabolic process	0.01
GO:0046496	nicotinamide nucleotide metabolic process	0.01
GO:0006740	NADPH regeneration	0.01
GO:0006739	NADP metabolic process	0.01
GO:0005996	monosaccharide metabolic process	0.01
GO:0008152	metabolic process	0.01
GO:0046483	heterocycle metabolic process	0.01
GO:0006091	generation of precursor metabolites and energy	0.01
GO:0051186	cofactor metabolic process	0.01
GO:0005975	carbohydrate metabolic process	0.01
GO:0016052	carbohydrate catabolic process	0.01
GO:0016051	carbohydrate biosynthetic process	0.01
GO:0009058	biosynthetic process	0.01

**Supplementary table 3. Gene ontology enrichment analysis of genes differentially alternatively spliced in *spt4R3* involved in cell division, development and biogenesis.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	cell division, development and biogenesis	p-value
GO:0015979	photosynthesis	1.50E-04
GO:0036211	protein modification process	0.01
GO:0070271	protein complex biogenesis	0.01
GO:0045893	positive regulation of transcription, DNA-dependent	0.01
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.01
GO:0051173	positive regulation of nitrogen compound metabolic process	0.01
GO:0010557	positive regulation of macromolecule biosynthetic process	0.01
GO:0010628	positive regulation of gene expression	0.01
GO:0043412	macromolecule modification	0.01
GO:0006464	cellular protein modification process	0.01
GO:0009987	cellular process	0.01
GO:0034641	cellular nitrogen compound metabolic process	0.01
GO:0006725	cellular aromatic compound metabolic process	0.01

**Supplementary table 4. Gene ontology enrichment analysis of genes differentially alternatively spliced in *spt4R3* involved in RNA splicing, RNA processing, Stress response and Light response.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	RNA splicing	p-value
GO:0000398	mRNA splicing, via spliceosome	0.000601
GO:0000377	RNA splicing, via transesterification reactions	0.0008
GO:0000375	RNA splicing, via transesterification reactions	0.0008

GO-term	RNA processing	p-value
GO:0051254	positive regulation of RNA metabolic process	0.01
GO:0006397	mRNA processing	0.01
GO:0016071	mRNA metabolic process	0.01

GO-term	Stress response	p-value
GO:0006950	response to stress	0.000833

GO-term	Light response	p-value
GO:0015979	photosynthesis	0.00015
GO:0019684	photosynthesis, light reaction	0.01

**Supplementary table 5. Gene ontology enrichment analysis of genes differentially alternatively spliced in *fflIs* involved in metabolic process.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	Metabolic processes	p-value
GO:0006007	glucose catabolic process	1.08E-07
GO:0046365	monosaccharide catabolic process	1.12E-07
GO:0019320	hexose catabolic process	1.12E-07
GO:0044724	single-organism carbohydrate catabolic process	1.33E-07
GO:0006006	glucose metabolic process	3.28E-07
GO:0019318	hexose metabolic process	5.92E-07
GO:0016052	carbohydrate catabolic process	1.12E-06
GO:0005996	monosaccharide metabolic process	1.55E-06
GO:0044237	cellular metabolic process	4.51E-06
GO:0071704	organic substance metabolic process	2.53E-05
GO:0008152	metabolic process	2.87E-05
GO:0006091	generation of precursor metabolites and energy	3.29E-05
GO:0044723	single-organism carbohydrate metabolic process	0.000112
GO:0044238	primary metabolic process	0.000201
GO:0006739	NADP metabolic process	0.000227
GO:1901564	organonitrogen compound metabolic process	0.000259
GO:0046496	nicotinamide nucleotide metabolic process	0.000272
GO:0019362	pyridine nucleotide metabolic process	0.000276
GO:0072524	pyridine-containing compound metabolic process	0.000317
GO:0044712	single-organism catabolic process	0.000399
GO:0006733	oxidoreduction coenzyme metabolic process	0.000655
GO:0044272	sulfur compound biosynthetic process	0.01
GO:0044281	small molecule metabolic process	0.01
GO:0044710	single-organism metabolic process	0.01
GO:0006793	phosphorus metabolic process	0.01
GO:0006796	phosphate-containing compound metabolic process	0.01
GO:0019637	organophosphate metabolic process	0.01
GO:1901575	organic substance catabolic process	0.01
GO:1901576	organic substance biosynthetic process	0.01
GO:0009117	nucleotide metabolic process	0.01
GO:0006753	nucleoside phosphate metabolic process	0.01
GO:0055086	nucleobase-containing small molecule metabolic process	0.01
GO:0006807	nitrogen compound metabolic process	0.01
GO:0043170	macromolecule metabolic process	0.01
GO:0051186	cofactor metabolic process	0.01
GO:0006732	coenzyme metabolic process	0.01
GO:0034641	cellular nitrogen compound metabolic process	0.01
GO:0044260	cellular macromolecule metabolic process	0.01
GO:0009056	catabolic process	0.01
GO:0005975	carbohydrate metabolic process	0.01
GO:0009058	biosynthetic process	0.01

**Supplementary table 6. Gene ontology enrichment analysis of genes differentially alternatively spliced in *tflls* involved in cell division, development and biogenesis.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	Cell division, development and biogenesis	p-value
GO:0006098	pentose-phosphate shunt	0.000203
GO:0006740	NADPH regeneration	0.000207
GO:0009657	plastid organization	0.000261
GO:0070271	protein complex biogenesis	0.000474
GO:0009987	cellular process	0.000773
GO:0044711	single-organism biosynthetic process	0.01
GO:0042254	ribosome biogenesis	0.01
GO:0022613	ribonucleoprotein complex biogenesis	0.01
GO:0071822	protein complex subunit organization	0.01
GO:0006461	protein complex assembly	0.01
GO:0009773	photosynthetic electron transport in photosystem I	0.01
GO:0009767	photosynthetic electron transport chain	0.01
GO:0006996	organelle organization	0.01
GO:0043933	macromolecular complex subunit organization	0.01
GO:0065003	macromolecular complex assembly	0.01
GO:0090391	granum assembly	0.01
GO:0006096	glycolysis	0.01
GO:0043623	cellular protein complex assembly	0.01
GO:0034622	cellular macromolecular complex assembly	0.01
GO:0071475	cellular hyperosmotic salinity response	0.01
GO:0071474	cellular hyperosmotic response	0.01
GO:0044085	cellular component biogenesis	0.01
GO:0022607	cellular component assembly	0.01
GO:0044249	cellular biosynthetic process	0.01

**Supplementary table 7. Gene ontology enrichment analysis of genes differentially alternatively spliced in *tflls* involved in RNA processing, Stress response and Light response.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	RNA processing	p-value
GO:0006364	rRNA processing	0.01
GO:0016072	rRNA metabolic process	0.01
GO:0034470	ncRNA processing	0.01
GO:0034660	ncRNA metabolic process	0.01

GO-term	Stress response	p-value
GO:0050896	response to stimulus	0.01
GO:0010035	response to inorganic substance	0.01
GO:0009628	response to abiotic stimulus	0.01



GO-term	Light response	p-value
GO:0015979	photosynthesis	3.09E-05
GO:0010207	photosystem II assembly	0.000119
GO:0019684	photosynthesis, light reaction	0.01

**Supplementary table 8a. Gene ontology enrichment analysis of genes differentially alternatively spliced in *iws1* involved in metabolic process.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function. All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	metabolic processes	p-value
GO:0006790	sulfur compound metabolic process	5.93E-05
GO:0044723	single-organism carbohydrate metabolic process	4.97E-08
GO:0005975	carbohydrate metabolic process	1.58E-07
GO:0044710	single-organism metabolic process	4.85E-07
GO:0006807	nitrogen compound metabolic process	6E-07
GO:0006006	glucose metabolic process	1.16E-06
GO:0005996	monosaccharide metabolic process	1.51E-06
GO:0006007	glucose catabolic process	2.08E-06
GO:0044711	single-organism biosynthetic process	2.14E-06
GO:0046365	monosaccharide catabolic process	2.16E-06
GO:0019320	hexose catabolic process	2.16E-06
GO:0019318	hexose metabolic process	2.6E-06
GO:0044724	single-organism carbohydrate catabolic process	2.69E-06
GO:0046483	heterocycle metabolic process	2.75E-06
GO:1901360	organic cyclic compound metabolic process	1.21E-05
GO:0006098	pentose-phosphate shunt	1.75E-05
GO:0006739	NADP metabolic process	2.11E-05
GO:0046496	nicotinamide nucleotide metabolic process	2.85E-05
GO:0019362	pyridine nucleotide metabolic process	2.94E-05
GO:0051186	cofactor metabolic process	3.09E-05
GO:0072524	pyridine-containing compound metabolic process	3.69E-05
GO:0016052	carbohydrate catabolic process	3.71E-05
GO:0071704	organic substance metabolic process	3.72E-05
GO:0044238	primary metabolic process	3.81E-05
GO:0008152	metabolic process	4.07E-05
GO:0006139	nucleobase-containing compound metabolic process	4.27E-05
GO:0044283	small molecule biosynthetic process	6.61E-05
GO:1901575	organic substance catabolic process	7.62E-05
GO:0006732	coenzyme metabolic process	9.65E-05
GO:0019637	organophosphate metabolic process	0.000113
GO:0006733	oxidoreduction coenzyme metabolic process	0.000123
GO:0019752	carboxylic acid metabolic process	0.000145
GO:1901564	organonitrogen compound metabolic process	0.000146
GO:0043436	oxoacid metabolic process	0.000194
GO:0006082	organic acid metabolic process	0.000196
GO:0016053	organic acid biosynthetic process	0.000222
GO:0046394	carboxylic acid biosynthetic process	0.000222
GO:0016051	carbohydrate biosynthetic process	0.000227
GO:0044699	single-organism process	0.000441
GO:0009058	biosynthetic process	0.00054
GO:1901576	organic substance biosynthetic process	0.000543
GO:0044281	small molecule metabolic process	0.000699

**Supplementary table 8b. Gene ontology enrichment analysis of genes differentially alternatively spliced in *iws1* involved in metabolic process.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function. All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	metabolic processes	p-value
GO:0044272	sulfur compound biosynthetic process	0.01
GO:0000096	sulfur amino acid metabolic process	0.01
GO:0000097	sulfur amino acid biosynthetic process	0.01
GO:0005982	starch metabolic process	0.01
GO:0019252	starch biosynthetic process	0.01
GO:0044802	single-organism membrane organization	0.01
GO:0044712	single-organism catabolic process	0.01
GO:0009069	serine family amino acid metabolic process	0.01
GO:0009070	serine family amino acid biosynthetic process	0.01
GO:1901968	regulation of polynucleotide 3'-phosphatase activity	0.01
GO:1901971	regulation of DNA-5-methylcytosine glycosylase activity	0.01
GO:0006090	pyruvate metabolic process	0.01
GO:1901969	positive regulation of polynucleotide 3'-phosphatase activity	0.01
GO:0010922	positive regulation of phosphatase activity	0.01
GO:0044093	positive regulation of molecular function	0.01
GO:1901972	positive regulation of DNA-5-methylcytosine glycosylase activity	0.01
GO:0043085	positive regulation of catalytic activity	0.01
GO:0009657	plastid organization	0.01
GO:0006793	phosphorus metabolic process	0.01
GO:0006644	phospholipid metabolic process	0.01
GO:0008654	phospholipid biosynthetic process	0.01
GO:0046471	phosphatidylglycerol metabolic process	0.01
GO:0006655	phosphatidylglycerol biosynthetic process	0.01
GO:0006796	phosphate-containing compound metabolic process	0.01
GO:0055114	oxidation-reduction process	0.01
GO:0090407	organophosphate biosynthetic process	0.01
GO:1901566	organonitrogen compound biosynthetic process	0.01
GO:0009311	oligosaccharide metabolic process	0.01
GO:0009117	nucleotide metabolic process	0.01
GO:0006753	nucleoside phosphate metabolic process	0.01
GO:0055086	nucleobase-containing small molecule metabolic process	0.01
GO:0090304	nucleic acid metabolic process	0.01
GO:0032787	monocarboxylic acid metabolic process	0.01
GO:0000023	maltose metabolic process	0.01
GO:0008610	lipid biosynthetic process	0.01

**Supplementary table 9. Gene ontology enrichment analysis of genes differentially alternatively spliced in *iws1* involved in cell division, development and biogenesis.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	Cell division,development and biogenesis	p-value
GO:0010028	xanthophyll cycle	0.01
GO:0005975	carbohydrate metabolic process	1.58E-07
GO:0044237	cellular metabolic process	4.71E-06
GO:0034641	cellular nitrogen compound metabolic process	8.97E-06
GO:0015979	photosynthesis	0.000013
GO:0006740	NADPH regeneration	0.000018
GO:0006725	cellular aromatic compound metabolic process	2.62E-05
GO:0009987	cellular process	3.18E-05
GO:0016052	carbohydrate catabolic process	3.71E-05
GO:0019752	carboxylic acid metabolic process	0.000145
GO:0009056	catabolic process	0.000184
GO:0046394	carboxylic acid biosynthetic process	0.000222
GO:0016051	carbohydrate biosynthetic process	0.000227
GO:0006081	cellular aldehyde metabolic process	0.000441
GO:0044249	cellular biosynthetic process	0.000481
GO:0009058	biosynthetic process	0.00054
GO:0006878	cellular copper ion homeostasis	0.000721
GO:0010207	photosystem II assembly	0.00097
GO:0042254	ribosome biogenesis	0.01
GO:0022613	ribonucleoprotein complex biogenesis	0.01
GO:0008104	protein localization	0.01
GO:0070271	protein complex biogenesis	0.01
GO:0009657	plastid organization	0.01
GO:0009767	photosynthetic electron transport chain	0.01
GO:0006827	high-affinity iron ion transport	0.01
GO:0033692	cellular polysaccharide biosynthetic process	0.01
GO:0044255	cellular lipid metabolic process	0.01
GO:0044085	cellular component biogenesis	0.01
GO:0034637	cellular carbohydrate biosynthetic process	0.01
GO:0008652	cellular amino acid biosynthetic process	0.01
GO:0015670	carbon dioxide transport	0.01
GO:1901135	carbohydrate derivative metabolic process	0.01

**Supplementary table 10. Gene ontology enrichment analysis of genes differentially alternatively spliced in *iws1* involved in RNA splicing, RNA processing, Stress response and Light response.** The Gene Ontology (GO) analysis was performed using *Plaza* (3.0 Dicots Workbench) to identify overrepresented groups in terms of biological function. All overrepresented GO terms with false discovery rate (FDR) < 0.05 are shown in the table.

GO-term	RNA splicing	p-value
GO:0000377	RNA splicing, via transesterification reaction	3.72E-05
GO:0000398	mRNA splicing, via spliceosome	2.42E-05
GO:0000375	RNA splicing, via transesterification reaction	3.72E-05
GO:0008380	RNA splicing	0.000489

GO-term	RNA processing	p-value
GO:0006364	rRNA processing	0.000542
GO:0006396	RNA processing	1.21E-05
GO:0016071	mRNA metabolic process	1.86E-05
GO:0006397	mRNA processing	0.000135
GO:0016072	rRNA metabolic process	0.000609
GO:0016070	RNA metabolic process	0.01
GO:0034470	ncRNA processing	0.01
GO:0034660	ncRNA metabolic process	0.01

GO-term	Stress response	p-value
GO:0006950	response to stress	0.01
GO:0009628	response to abiotic stimulus	0.000526
GO:0050896	response to stimulus	0.01
GO:0009651	response to salt stress	0.01
GO:0010114	response to red light	0.01
GO:0006970	response to osmotic stress	0.01
GO:0009642	response to light intensity	0.01
GO:0010035	response to inorganic substance	0.01
GO:0010218	response to far red light	0.01
GO:0046686	response to cadmium ion	0.01
GO:0009637	response to blue light	0.01

GO-term	Light stimulus	p-value
GO:0010114	response to red light	0.01
GO:0071486	cellular response to high light intensity	0.000516
GO:0009642	response to light intensity	0.01
GO:0010218	response to far red light	0.01
GO:0009637	response to blue light	0.01
GO:0034644	cellular response to UV	0.01
GO:0071484	cellular response to light intensity	0.01

**Supplementary Table 11 a.** Alternative splice forms were co-amplified and quantified using a Bioanalyzer. Ratios were normalized to the corresponding control Col-0. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS

## At3g58510

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	1.322709163	3.824701195	5.019920319	1.673306773
1	1.530120482	4.128514056	4.979919679	1.831325301
1	1.480225989	3.95480226	5.81920904	2.751412429

## At1g68490

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	1.139784946	0.636200717	0.367383513	0.630824373
1	0.935281837	0.549060543	0.417536534	0.668058455
1	0.885714286	0.567142857	0.282857143	0.761428571

## At3g04500

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	1.05366162	2.09596	1.755050505	1.325757576
1	0.57202797	1.72028	2.321678322	1.388811189
1	1.18239437	2.535211	2.169014085	1.718309859
1	0.93602798	2.11715	2.081914304	1.477626208

## At3g61860: RS31

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	1.465263158	4.8	5.515789474	2.92631579
1	1.485087719	4.684210526	4.684210526	2.35877193
1	1.604761905	5.654761905	6.011904762	3.86904762

## At2g25970

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	0.771929825	1.842105263	1.84210526	1.210526316
1	1.666666667	3.5	2.333333333	0.283333333
1	1.222222222	1.769444444	1.572222222	0.908333333

## At1g02840: SR34

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	spt4R3
1	1.24137931	2.103448	4.551724138	2.172413793
1	0.73	2.516667	3.666666667	1.666666667
1	0.608695652	2.608696	1.652173913	1.652173913

**Supplementary Table 11 b.** Alternative splice forms were co-amplified and quantified using a Bioanalyzer. Ratios were normalized to the corresponding control Col-0. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS

## At3g22210

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	<i>spt4R3</i>
1	1	0.666666667	0.533333333	0.4
1	0.916666667	0.666666667	0.666666667	0.75
1	0.75	0.75	0.9375	0.5625
1	0.866666667	0.8	0.666666667	0.6

## At1g50902

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	<i>spt4R3</i>
1	0.983051	1.025424	0.813559	0.288136
1	1.144578	0.783133	0.674699	0.710843
1	0.945946	0.801802	0.756757	0.306306

## At4g38280

Col-0	<i>tfls</i>	<i>iws1</i>	<i>tfls/iws1</i>	<i>spt4R3</i>
1	1.456521739	0.402173913	0.367391304	0.782608696
1	0.152777778	0.179166667	0.294444444	0.444444444
1	0.795	0.44	0.85	1
1	0.842105263	0.426315789	0.947368421	0.789473684

## At2g04039

Col-0	<i>tfls</i>	<i>iws</i>	<i>tfls/iws1</i>	<i>spt4R3</i>
1	0.170068	0.646259	1.142857	0.707483
1	0.534884	2.523256	0.638372	1.965116
1	1.333333	1.583333	1.166667	2.166667

**Supplementary table 12. ChIP seq sample information.** The specific KFB ID, the index primer associated with the ChIP antibody and respective replicates. The DNA input measured before the ChIP experiment with the Qubit and the KAPA PCR performed in the KFB.

Sequencing library database							
KFB ID	Index Primer#	line	ChIP antibody	DNA input	PCR cycles	QUBIT	KAPA (nM)
					cycler		
S314b_01	1	WT IV repl. (wt - Spt4R3)	I	15,8 ng/ul	14	0,614	559
S314b_02	2	WT IV repl. (wt - Spt4R3)	H3	0,203	14	0,063	251
S314b_03	3	WT IV repl. (wt - Spt4R3)	H3 K36me3	0,464	14	0,150	799
S314b_04	4	WT IV repl. (wt - Spt4R3)	S2P	0,0812	14	0,118	187
S314b_05	5	WT IV repl. (wt - Spt4R3)	S5P	0,117	14	2,58	200
S314b_06	6	Spt4R3 IV repl. (wt - Spt4R3)	I	too high	14	0,069	659
S314b_07	7	Spt4R3 IV repl. (wt - Spt4R3)	H3	1,98	14	0,434	278
S314b_08	8	Spt4R3 IV repl. (wt - Spt4R3)	H3 K36me3	0,68	14	0,856	758
S314b_09	9	Spt4R3 IV repl. (wt - Spt4R3)	S2P	0,032	14	0,504	13
S314b_10	10	Spt4R3 IV repl. (wt - Spt4R3)	S5P	0,76	14		88
S314b_11	11	WT VI repl. (wt - Spt4R3)	H3	1,26	14	0,198	670
S314b_12	12	WT VI repl. (wt - Spt4R3)	H3 K36me3	0,572	14	0,638	113
S314b_13	13	WT VI repl. (wt - Spt4R3)	S2P	0,0644	14		62
S314b_14	14	WT VI repl. (wt - Spt4R3)	S5P	0,058	14		44
S314b_15	15	lws1a IV repl (wt-iws1a)	I	13,5 ng/ul	14	0,168	970
S314b_16	16	lws1a IV repl (wt-iws1a)	H3	0,279	14	11,4	476
S314b_18	18	lws1a IV repl (wt-iws1a)	H3 K36me3	0,936	14	0,162	683
S314b_19	19	lws1a IV repl (wt-iws1a)	S2P	0,0236	14	0,158	48
S314b_20	20	lws1a IV repl (wt-iws1a)	S5P	0,028	14	0,282	37
S314b_21	21	Spt4R3 VI repl. (wt-Spt4R3)	H3	0,66	14	5,52	632
S314b_22	22	Spt4R3 VI repl. (wt-Spt4R3)	H3 K36me3	0,0964	14	4,42	110
S314b_23	23	Spt4R3 VI repl. (wt-Spt4R3)	S2P	0,0784	14	0,160	826
S314b_25	25	Spt4R3 VI repl. (wt-Spt4R3)	S5P	0,343	14	0,640	58
S314b_27	27	lws1a I repl. (wt-iws1a)	H3	0,21	14		594
S314b_17	17	lws1a I repl. (wt-iws1a)	H3 K36me3	0,278	14		459
S314b_24	24	lws1a I repl. (wt-iws1a)	S2P	0,022	14	0,404	91
S314b_26	26	lws1a I repl. (wt-iws1a)	S5P	0,32	14		50
S314b_09-2	9-2						40
S314b_24-2	24-2						33
S314b_25-2	25-2						26

**Supplementary table 13 Phenotypic analysis of *mrg1/mrg2/iws1* compared to the double mutant *mrg1/mrg2*, single mutant *iws1* and Col-0.** All data are means  $\pm$ SD base on n. =15 individuals. The statistical analysis were made according a multi comparisons Tukey's test (p-value < 0.05) analysis made with *Past3* (Hammer et al., 2001).

Genotype	Bolting [DAS]	Rosette diameter at bolting [mm]	no. of leaves at bolting	Rosette Diameter 28 DAS [mm]
Col-0	22.78 $\pm$ 0.43	76.94 $\pm$ 12.38	10.89 $\pm$ 1.37	93.72 $\pm$ 24.14
<i>iws1</i>	21.76 $\pm$ 0.75	27.06 $\pm$ 3.09	6.65 $\pm$ 1.37	41.76 $\pm$ 7.28
<i>mrg1/mrg2</i>	26.88 $\pm$ 1.22	97.65 $\pm$ 8.50	12.71 $\pm$ 1.61	102.94 $\pm$ 15.11
<i>mrg1/mrg2/iws1</i>	29.38 $\pm$ 0.51	61.15 $\pm$ 10.03	13.23 $\pm$ 1.17	47.7 $\pm$ 15.67

**Supplementary table 14 Phenotypic analysis of *mrg1/mrg2/iws1* compared to the double mutant *mrg1/mrg2*, single mutant *iws1* and Col-0.** All data are means  $\pm$ SD base on n. =15 individuals. The statistical analysis were made according a multi comparisons Tukey's test (p-value < 0.05) analysis made with *Past3* (Hammer et al., 2001).

Genotype	Plant height 28 DAS [mm]	Plant height 42 DAS [cm]	Primary infl. 42 DAS	Primary infl. 50 DAS	Plant height 50 DAS [cm]
Col-0	39.41 $\pm$ 25.97	28.78 $\pm$ 0.73	6.44 $\pm$ 0.62		
<i>iws1</i>	46.18 $\pm$ 14.85	14.88 $\pm$ 0.78	3.82 $\pm$ 0.81	6.13 $\pm$ 0.34	22.19 $\pm$ 3.02
<i>mrg1/mrg2</i>	44.09 $\pm$ 31.84	25.76 $\pm$ 2.28	6.12 $\pm$ 0.70	7.35 $\pm$ 0.61	27.35 $\pm$ 1.87
<i>mrg1/mrg2/iws1</i>		8.85 $\pm$ 0.90	1.00	6.08 $\pm$ 0.64	23.92 $\pm$ 2.47

**Supplementary table 15. Transcription-related proteins co-purifying with PTB2-SG.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

PTB2-SG	AGI	PROCESS	COMPLEX	INTERACTOR
196/3	AT4G08350	transcription	SPT4/SPT5	SPT5-2
253/2	AT1G32130	transcription	IWS1	IWS1a
308/2	AT5G10270	transcription	P-TEFb	CDKC;1
206/2	AT5G45190	transcription	P-TEFb	CYCT1;5
212/2	AT4G21710	transcription	RNAPII	NRPB2
240/2	AT4G35800	transcription	RNAPII	NRPB1
228/3	AT3G22590	transcription	PAF-C	CDC73
128/2	AT2G06210	transcription	PAF-C	CTR9, ELF8, VIP6
199/3	AT4G29830	transcription	PAF-C	SKI8, VIP3
293/2	AT1G79730	transcription	PAF-C	PAF1, ELF7
156/2	AT5G61150	transcription	PAF-C	LEO1, VIP4

**Supplementary table 16. Transcription-related proteins co-purifying with GS-MRG1** List of transcription-related proteins co-immunoprecipitated with MRG1-SG after data proceeding. The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

GS-MRG1	AGI	PROCESS	COMPLEX	INTERACTOR
634/3	AT3G24870	Transcription	SWR1/NuA4	AtEAF1B
841/2	AT2G19480	Transcription	putative histone chaperones	AtNAPL2
480/2	AT4G29830	Transcription	PAF-C	SKI8, VIP3
256/2	AT5G45600	Transcription	SWR1/NuA4	AtYAF9A
204/2	AT2G19540	Transcription	putative histone chaperones	AtCAF1CL6
167/2	AT4G14385	Transcription	SWR1/NuA4	EAF6

**Supplementary table 17. Transcription-related proteins co-purifying with GS-MRG2.** List of transcription-related proteins co-immunoprecipitated with GS-MRG2 after data proceeding. The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

GS-MRG2	AGI	PROCESS	COMPLEX	INTERACTOR
226/2	AT4G21710	transcription	RNAPII	NRPB2
669/2	AT2G06210	transcription	PAF-C	CTR9, ELF8
500/2	AT3G22590	transcription	PAF-C	CDC73
144/2	AT1G79730	transcription	PAF-C	PAF1, ELF7
382/2	AT1G07660	transcription	Histone H4 family	Histone H4



**Supplementary table 18a. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o	
						benzonase	PTB2-SG
AT5G19300	Mtase,C9orf114	1003/3		409/2		845/2	579/3
AT2G44150	Mtase,ASHH3	525/3	801/3		1003/2	511/2	
AT1G02740	SWR1/NuA4,MRG2	2198/3	2403/4		200/2	3044/2	
AT4G37280	SWR1/NuA4,MRG1			2794/3	2378/4		
AT2G43770	U5 snRNP,atU5-40 Prp8	475/2		337/3		421/2	1068/3
AT2G33730	U5 snRNP,atU5-100KD	571/2				327/2	537/3
AT5G16780	U4/U6.U5 tri snRNP,MDF	682/3		1058/3		1403/2	1902/3
AT3G05760	U4/U6.U5 tri snRNP,U4/U6.U5			448/2		441/2	458/2
AT1G04080	U1 snRNP,atPrp39a late flowering					133/2	
AT3G61860	SR proteins,atRSp31					146/2	890/3
AT1G56290	NTC-associated,CwfJ-like protein			422/3		463/2	517/2
AT2G38770	NTC-associated,atAquarius EMB2765			795/2		241/2	3468/3
AT1G07360	NTC-associated,atECM2-1a	224/3				180/2	778/3
AT2G27100	splicing,SERRATE		1359/2	535/2		920/2	966/3
AT5G63120	mRNA binding , RNA helicase 30			391/3		170/2	1081/3
AT3G26420	hnRNP family,ATRZ-1A					388/2	106/2
AT1G50300	hnRNP family,at-hnRNP-P					113/2	
AT3G58570	EJC/mRNP, RNA helicase 52		292/2	365/3		360/2	778/3
AT1G77180	Core NTC,SKIP			502/2		546/2	1688/3
AT2G33340	Core NTC, (MAC3B), PRP19B	474/2				139/2	1631/3
AT3G55200	17S U2 snRNP,atSAP130a	876/2		633/2		193/2	2066/3
AT2G18510	17S U2 snRNP	237/3		328/3	176/3	135/2	492/3
AT5G12190	17S U2 snRNP,atP14-1					134/2	384/2
AT1G20960	U5 snRNP,atU5	2044/2		1570/2			6264/3
AT4G03430	U5 snRNP,atU5	555/2					1246/3
AT5G09390	U5 snRNP,CD2 antigen	135/2					
AT2G40650	U4/U6.U5 tri snRNP,splicing	177/2					623/3
AT1G28060	U4/U6 snRNP,atSAP90-1	399/3					639/3
AT2G41500	U4/U6 snRNP,atSAP60	392/2					586/3
AT5G51410	U1 snRNP,atLuc7-rl	133/2					354/3
AT3G49430	SR proteins,AtSR34a	387/2					934/3
AT5G64200	SR proteins,atSC35	98/2					164/3
AT1G09770	Core NTC,CDC5	753/2		377/2			2809/3
AT5G28740	Core NTC,splicing	640/2					1653/3
AT1G04510	Core NTC,MOS4	410/2					1346/3
AT3G18790	Core NTC,atlsy1	393/2					931/3
AT4G15900	Core NTC,PRL1	324/2					959/3
AT5G41770	Core NTC,atCRN1c	287/2					755/3
AT4G21660	17S U2 snRNP,atSF3b150	451/2		420/3			1155/3
AT1G14650	17S U2 snRNP,atSAP114-1a	362/2					1570/3
AT5G25060	17S U2 associated,atSR140-1 U2	711/2		502/3			1658/3
AT5G53180	PTB2						2363/3
AT5G61140	U5 snRNP,						321/2
AT5G08290	U5 snRNP						121/2
AT5G37370	U4/U6.U5						479/2
AT1G60170	U4/U6						634/3
AT2G38730	U4/U6 snRNP,atTri-20						463/2
AT1G60200	U1snRNP related,RNA-binding protein						737/2
AT1G44910	U1snRNP related,atPRP40A						451/2
AT3G07860	U11/U12 oprotein						186/2
AT3G50670	U1 snRNP,atU1-70K						1738/3
AT5G17440	U1 snRNP,atLuc7b						896/3
AT3G03340	U1 snRNP,atLuc7a						826/3
AT2G47580	U1 snRNP,atU1A						843/2
AT4G03120	U1 snRNP						385/2
AT1G03910	SR-related proteins						315/3
AT2G29210	SR-related proteins						246/2
AT2G46610	SR proteins,atRSp32						661/3
AT1G55310	SR proteins,atSR33/atSCL33						566/3
AT3G55460	SR proteins,						574/3
AT4G31580	SR proteins,atRSZp22/atSRZ22						556/3

**Supplementary table 18b. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT1G23860	SR proteins,atRSzp21/atSRZ21						459/3
AT5G18810	SR proteins,atSCL28						339/3
AT3G13570	SR proteins,AtSCL30a						430/3
AT4G35785	SR proteins						452/3
AT1G02840	SR proteins,SR1, ATSR34						484/3
AT2G24590	SR proteins,atRSzp22a						354/3
AT1G09140	SR proteins,ATSR30						544/3
AT2G37340	SR proteins,atRSZ33						338/3
AT5G52040	SR proteins,atRSp41						556/2
AT3G53500	SR proteins,RSZ32						156/2
AT3G44850	SR protein kinase,atSRPK2c						777/3
AT5G22840	SR protein kinase,atSRPK2a						432/3
AT4G32660	SR protein kinase						401/3
AT4G24740	SR protein kinase,						450/3
AT4G36690	Splice site selection,atU2AF65a						689/3
AT1G60900	Splice site selection,atU2AF65b						763/3
AT5G42820	Splice site selection,atU2AF35						371/2
AT1G27650	Splice site selection,atU2AF35a						299/2
AT1G32490	NTC-associated,atPrp2a ESP3						1403/3
AT1G80930	NTC-CWC22						1037/3
AT1G10580	NTC-associated,atPrp17-1						661/3
AT2G29580	NTC-associated,atECM2-1b						527/3
AT3G05070	NTC-associated,cwf18						271/3
AT3G13200	NTC-associated,CWC15, EMB2769						238/3
AT3G23900	NTC-associated,atSRM300-like						424/3
AT3G57910	NTC-associated,D111/G						176/3
AT2G36130	NTC-associated,						592/2
AT4G33060	NTC-associated,CWC27,						582/2
AT1G33520	NTC-associated,MOS2						442/2
AT4G21110	NTC-associated						264/2
AT3G02710	NTC-associated						615/2
AT1G25682	NTC-associated						210/2
AT1G01350	NTC-associated						131/2
AT4G16280	hnRNP family,FCA ALPHA						1040/3
AT2G44710	hnRNP family,hnRNP-R3						1181/3
AT1G60650	hnRNP family,at-hnRNP-G3						532/3
AT5G04280	hnRNP family,at-hnRNP-G1 RNA						375/3
AT3G19130	hnRNP family,AtRBP47b						365/3
AT5G54900	hnRNP family,ATRBP45A						524/3
AT1G47500	hnRNP family,AtRBP47c'						232/2
AT3G15010	hnRNP family,UBA2c RNA	105/2					499/2
AT5G40490	hnRNP family,AtRNPA/B_8a						397/2
AT3G13224	hnRNP family,AtRNP_N1						425/2
AT2G45640	EJC/mRNP,HDAC subunit SAP18						569/3
AT1G16610	EJC/mRNP,atSR45/atRNPS1						267/3
AT1G02140	EJC/mRNP,HAP1;						840/2
AT1G51510	EJC/mRNP,Y14,						187/2
AT5G64270	17S U2 snRNP,atSAP155			987/3			3076/3
AT5G06160	17S U2 snRNP,atSAP61 s			117/2			664/3
AT1G09760	17S U2 snRNP,atU2A'						376/3
AT2G32600	17S U2 snRNP						476/2
AT2G30260	17S U2 snRNP,atU2B''b						158/2
AT1G30480	trafficking protein particle complex						542/3
AT2G47330	17S U2 associated						664/2
AT5G23300	dihydroorotate dehydrogenase	524/3				655/2	577/2
AT4G22690	cytochrome P450	521/3					
AT3G25860	dihydrolipoamide S-acetyltransferase	443/3					
AT2G47210	AtSWC4	439/3	460/3			617/2	
AT2G20280	Zinc finger	410/3				183/2	243/2
AT2G22475	Encodes GL2-expression modulator	390/3				463/2	240/2
AT2G42710	Ribosomal protein L1p	384/3		490/3	431/3	410/2	645/3
AT5G46750	GAP domain (AGD)	342/3	207/2		147/2	416/2	
AT1G34430	2-oxoacid	285/3				235/2	305/3
AT2G47960	trafficking protein particle complex	282/3				183/2	

**Supplementary table 18c. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT5G13020	Emsy	281/3					119/2
AT4G31810	ATP-dependent	271/3		561/3	382/4	224/2	
AT3G15090	GroES-like zinc-binding	269/3				444/2	
AT5G66680	SOT48	250/3				356/2	284/3
AT1G12840	Encodes subunit C of the vacuolar	231/3				215/2	
AT5G23200	C5orf35	226/3					131/2
AT1G73840	CstF64 f	221/3		269/2		108/2	324/3
AT5G18420	CCR4-NOT	219/3	123/2			187/2	258/2
AT2G40660	Nucleic acid-binding	218/3				307/2	
AT2G36070	mitochondrial translocase complex.	207/3					173/3
AT5G66760	mitochondrial dehydrogenase	196/3				293/2	165/2
AT5G50340	DNA repair protein RadA-like protein	193/3					
AT1G21080	DNAJ h	187/3				338/2	104/2
AT1G29220	transcriptional regulator family protein	174/3					248/3
AT5G20200	nucleoporin-like	162/3				177/2	186/3
AT4G34660	SH3	155/3				144/	159/3
AT4G17950	AT hook motif DNA-binding protein	152/3	523/2			648/2	410/2
AT1G16870	mitochondrial 28S	149/3					215/2
AT5G58230	AtCAF1CL1	149/3					
AT4G17890	putative ADP-ribosylation factor G	121/3					
AT1G52570	phospholipase D alpha 2	1323/2		388/3		942/2	578/3
AT1G62390	36 carboxylate clamp (CC)	966/2		636/2		339/2	283/2
AT1G63810	nucleolar protein	917/2					149/2
AT3G21540	transducin/WD40	883/2					328/2
AT5G06970	Munc13-like	854/2					436/3
AT4G18060	SH3	751/2				176/2	438/3
AT1G31970	DEA(D/H)-box	702/2				167/2	418/2
AT2G06210	CTR9,	669/2					128/2
AT1G67680	SRP72	654/2		228/3		702/2	
AT4G24680	Encodes MOS1 (MODIFIER OF snc1)	643/2		1838/3		584/2	1487/3
AT4G39680	SAP	616/2		407/2	247/2	233/2	2559/3
AT3G08943	armadillo/beta-catenin-like	615/2			340/2		200/2
AT1G72150	novel cell-plate-associated protein	597/2				710/2	257/2
AT5G58070	TIL1	568/2		263		723/2	428/2
AT3G55410	2-oxoglutarate	563/2		293/2		615/2	827/3
AT3G62310	atPrp43-2a	526/2		727/2		286/2	1663/3
AT3G22590	CDC73	500/2					229/3
AT1G15200	protein-protein interaction r	486/2					1627/3
AT2G43970	RNA-binding protein	468/2		1054/2			679/2
AT3G46740	translocon outer membrane (TOC)	465/2					198/3
AT5G16840	binding partner of acd11	453/2		162/2		830/2	188/3
AT5G01380	trihelix transcription factor GT-3a	453/2					390/3
AT4G18465	ATP-dependent RNA helicase DDX35	452/2					898/3
AT2G03510	SPFH/Band	446/2					307/3
AT2G17200	Encodes a ubiquitin receptor	444/2				238/2	1147/2
AT4G21800	Encodes QQT2.	442/2	235/2			218/2	
AT2G14120	Encodes a dynamin related protein	440/2				487/2	258/3
AT4G21100	damaged DNA binding protein	432/2				470/2	149/2
AT5G13110	glucose-6-phosphate dehydrogenase	421/2				248/2	396/3
AT2G32920	protein disulfide-isomerase like 2-3	407/2					
AT5G16750	WD-40	402/2					
AT3G24870	AtEAF1B	402/2	329/2	634/3	513/2	314/2	
AT1G07660	Histone H4,Histone H4	382/2	233/3		180/3		
AT1G04170	factor eIF2 gamma	376/2				127/2	
AT3G12790	unknown protein	372/2					214/2
AT5G08450	Interacts with HDA6 and HDA19	370/2				338/2	189/2
AT1G60070	Adaptor protein complex AP-1	366/2				206/2	
AT1G07670	endomembrane-type CA-ATPase 4	358/2				166/2	
AT1G77470	,Replication factor	352/2				331/2	
AT2G01690	ARM repeat superfamily protein	339/2				349/2	239/2
AT4G02570	SCF ubiquitin ligase complexes	337/2				347/2	208/2
AT2G33040	gamma subunit of Mt ATP synthase	334/2				486/2	198/3
AT5G61970	SRP-like protein	332/2				304/2	
AT4G05420	DNA binding proteins	331/2				477/2	
AT1G55460	DNA/RNA-binding	331/2					
AT2G02090	SNF2	330/2				314/2	

**Supplementary table 18d. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-SG.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT1G47550	e exocyst complex	330/2				128/2	142/2
AT3G09350	Hsp70-binding protein 1	330/2	138/2			329/2	
AT5G22770	alpha-adaptin	327/2				342/2	109/2
AT2G17980	SLY1 Gene Family	326/2		174/2		418/2	161/2
AT5G49555	FAD/NAD(P)-binding	324/2				332/2	169/2
AT4G33650	dynamain superfamily	323/2				431/2	279/3
AT4G23850	AMP-dependent	322/2				147/2	
AT5G16620	chloroplast	322/2					
AT2G45520	coiled-coil	321/2		214/2		276/2	265/3
AT1G55890	Tetratricopeptide	320/2				237/2	599/2
AT5G49830	exocyst complex gene family	317/2				113/2	118/2
AT2G07050	biosynthesis of brassinosteroids	311/2				286/2	
AT5G06140	yeast retromer subunit VPS5	311/2				181/2	230/2
AT5G30510	small subunit ribosomal protein S1	306/2					258/2
AT2G21870	F1F0-ATP synthase	305/2				194/2	375/2
AT5G65750	2-oxoglutarate dehydrogenase,	304/2				381/2	
AT1G48900	SRP54 subunit protein	299/2				504/2	
AT3G57980	DNA-binding bromodomain-containing	297/2		367/3	298/3		
AT1G72320	(APUM)	292/2				114/2	
AT5G40770	prohibitin 3,prohibitin 3	289/2				343/2	354/2
AT3G16830	TOPLESS-related	288/2				147/2	185/2
AT1G05520	T25N20.17	286/2					
AT1G44900	minichromosome	284/2					
AT1G69070	nucleolar-like	277/2					174/3
AT4G02350	exocyst complex gene family	275/2				272/2	200/2
AT5G62270	ribosomal protein L20	273/2					263/3
AT3G02200	Proteasome component	269/2	221/2	829/3	922/3	376/2	165/2
AT4G36680	Tetratricopeptide repeat (TPR)-like	264/2				206/2	
AT2G22660	DUF1399 domain	263/2					
AT5G26760	Encodes RPAP2 IYO Mate	262/2					306/2
AT3G53710	GAP domain (AGD)	259/2				194/2	
AT4G24840	oligomeric	258/2				250/2	
AT1G60770	Tetratricopeptide	256/2					
AT4G11150	H+-ATPase subunit E isoform 1	255/2				339/2	
AT2G39810	EyIs/HOS1	254/2		338/2		290/2	154/2
AT5G13300	GTPase gene family	254/2					
AT1G09270	VirD2 and VirE2	251/2	172/2		401/2	212/2	218/3
AT4G17520	Ath2	249/2	212		375/3		
AT3G48860	coiled-coil	245/2		386/2		157/2	337/3
AT3G06480	DEAD	244/2					1115/3
AT5G10910	mraW	241/2					
AT5G10730	NAD(P)-binding	239/2				264/2	
AT3G07300	EIF-2B beta subunit	237/2					
AT5G11390	putative protein	230/2					
AT1G67930	Golgi transport complex protein	228/2				150/2	
AT4G21710	NRPB2	227/2	133/2				212/2
AT1G77250	RING/FYVE/PHD-type zinc finger	223/2	767/2	272/3		208/2	
AT2G38040	acetyl-CoA carboxylase	221/2				206/2	287/2
AT3G19590	BUB3.1	221/2					
AT5G25270	Ubiquitin-like	215/2					
AT5G39410	Saccharopine	214/2				280/2	
AT5G14580	,polyribonucleotide	211/2					133/2
AT3G06850	dihydrolipoamide	211/2				326/2	429/2
AT1G31440	SH3	207/2				148/2	169/2
AT1G69830	amylase	207/2				320/2	
AT1G20330	sterol-C24-methyltransferases	206/2				321/2	145/2
AT2G31810	ACT	205/2				513/2	148/2
AT2G27200	P-loop	205/2					316/2
AT5G67500	voltage-dependent anion channel	203/2				212/2	
AT3G09300	OSBP	202/2					
AT4G15020	hAT	201/2					214/3
AT4G31200	SWAP	198/2		441/3			701/3
AT1G67440	Minichromosome	196/2					229/3
AT3G14120	,Nup107	196/2					
AT1G19520	pentatricopeptide	193/2			392/2		285/2
AT1G30580	,GTP-binding	192/2					

**Supplementary table 18e. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT5G55670	RNA recognition pretein	190/2					576/3
AT2G29560	phosphoenolpyruvate enolase	186/2					
AT4G24550	AP4 complex	185/2				180/2	
AT1G53165	Protein kinase	182/2				161/2	306/2
AT5G09840	endonuclease or glycosyl hydrolase	181/2					421/3
AT1G30680	T5I8.13	180/2					358/2
AT5G03540	AtEXO70A1	187/2		237/2			172/2
AT3G04600	tryptophanyl-tRNA	177/2					
AT5G27540	GTPases	174/2					
AT1G53645	hydroxyproline-rich glycoprotein	174/2					108/2
AT5G13310	uncharacterized protein	174/2					
AT1G75660	XRN3 5'-3' exoribonuclease 3	172/2					487/2
AT5G41950	Tetratricopeptide repeat (TPR)-like	171/2				217/2	159/2
AT5G46280	DNA replication licensing factor	169/2					
AT5G27600	acyl-CoA synthetase	167/2				121/2	
AT1G01300	aspartyl protease family protein	167/2		402/3		325/2	346/2
AT4G33945	ARM	163/2				162/2	
AT1G48570	zinc finger (Ran-binding) family protein	162/2					277/3
AT3G62240	RING/U-box superfamily protein	160/2					
AT1G51570	Calcium-dependent lipid-binding	159/2					
AT1G54460	TPX2	158/2					228/2
AT2G36200	P-loop containing nucleoside	156/2					89/2
AT3G61820	Eukaryotic aspartyl protease	156/2				190/2	
AT3G45190	SIT4 phosphatase	156/2					
AT1G29030	Hypothetical protein	155/2					302/2
AT1G34360	translation initiation factor IF-3	155/2					202/2
AT2G22300	CAM binding transcription factor	154/2					322
AT1G68370	DnaJ-like protein	151/2				189/2	
AT3G48530	,SNF1-related protein kinase	146/2					
AT1G79730	PAF1, ELF7	144/2					293/2
AT3G54470	phosphoribosyltransferase	143/2					161/2
AT1G12920	peptide chain release factor sub. 1-2	142/2					
AT2G04350	long chain acyl-CoA synthetase 8	141/2				113/2	
AT5G05990	Mitochondrial	139/2					
AT1G80680	Nup96	138/2					
AT1G72560	LOS1/XPOT, export of tRNA	135/2				160/2	
AT1G52980	AT1G52980	134/2					
AT1G69250	NTF2-Like 10	131/2		436/2			
AT2G37230	(TPR)-like superfamily protein	131/2					716/2
AT4G01400	uncharacterized protein	130/2					167/3
AT1G74960	beta-ketoacyl-ACP synthase II	129/2					
AT3G20650	guanine-N7 methyltransferase 1	129/2					
AT1G26750	hypothetical protein	129/2			441/2	262/2	168/2
AT1G22730	MA3	128/2				118/2	
AT2G18720	eIF2S3/eIF-2-gamma	125/2					
AT1G62120	F19K23.6 gene product	124/2					
AT2G01410	NHL	123/2					
AT2G05920	Subtilase family protein	121/2					
AT4G01040	,A_IG002N01.31 gene product	119/2					158/2
AT1G51580	RNA-binding KH	118/2					244/3
AT2G39795	unknown	117/2					
AT1G03140	Prp18-like protein	116/2					504/3
AT1G48160	SRP19	114/2				196/2	
AT4G11260	SCF(TIR1)	113/2					
AT5G15550	WD-40 repeat family protein	109/2					
AT5G12330	LATERAL ROOT PRIMORDIUM 1	108/2					288/2
AT4G34490	CYCLASE ASSOCIATED PROTEIN	106/2					
AT2G17510	EMB2763	103/2				179/2	
AT5G15980	Pentatricopeptide repeat (PPR)	101/2					
AT3G20630	ubiquitin carboxyl-terminal hydrolase	99/2					
AT1G17130	uncharacterized protein	95/2					356/3
AT1G60620	NRPC3	95/2					
AT5G14460	Pseudouridine synthase family protein	95/2					228/2
AT5G56630	6-phosphofructokinase 7	91/2					342/3
AT3G23910	uncharacterized protein	90/»					
AT1G12470	zinc ion binding protein	88/2					

**Supplementary table 18f. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o	
						benzonase	PTB2-SG
AT5G02500	HSC70-1 ,heat shock protein 70		2558/4		1210/4		
AT5G09590	heat shock protein 70 (Hsc70-5)		1549/4		1081/4		
AT3G09440	protein heat shock protein 70-3		2148/4		903/4		
AT3G12580	HSP70 heat shock protein 70-4		2023/4		1087/3		
AT5G09810	mRNA is cell-to-cell mobile		1041/4		625/4		
AT2G29550	tubulin beta		1178/4				
AT3G60110	DNA-binding bromodomain		268/4		297/4	357/2	
AT4G20360	RAB GTPase homolog E1B		915/4		482/3		
AT1G26470	EAF7,unknown protein		460/4	1148/3	1018/4	883/2	
AT1G72370	acidic protein		218/4		528/4		
AT2G44060	Late embryogenesis abundant protein		298/4		278/2		
AT4G37910	mitochondrial heat shock protein 70-1		2259/3		1357/4		
AT5G49910	mRNA is cell-to-cell mobile		1093/3		1120/3		
AT5G02490	HSC70-2,heat shock protein 70		2318/3		1072/2		
AT1G48920	ATNUC-L1 (NUCLEOLIN LIKE 1)		401/3				
AT5G28540	HSP70 family		741/3		295/2		
AT5G43960	RNA recognition motif		305/3				
AT3G54110	PUMP2 family		188/3			271/2	298/2
AT4G20890	AT4G20890		1068/3		959/2		
AT4G01100	adenine nucleotide transporter 1		196/3			191/2	243/3
AT5G62690	tubulin beta-2/beta-3 chain		1603/3		646/4		
AT4G14960	tubulin alpha-6 chain		1021/3		590/4		
AT5G44340	beta tubulin gene,tubulin beta-4 chain		1645/3		891/3		
AT5G23860	beta-tubulin		1456/3				
AT5G19770	tubulin 3,tubulin alpha-3/alpha-5 chain		609/3		541/2		
AT3G13920	translational initiation factor 4A-1		640/3		388/2		
AT3G63460	WD-40		652/3		155/2		
AT2G36460	Aldolase superfamily protein		364/3		398/2		
AT5G08670	ATP synthase beta-subunit		362/3				
AT1G18450	AtARP4,unknown		337/3		456/3		
AT2G39730	Rubisco activase		569/3		245/2		
AT3G44110	DNAJ protein from E coli		611/3		235/4		
AT1G79530	chloroplast/plastid localized GAPDH		305/3		127/2		
AT3G52930	Aldolase superfamily protein		286/3		262/2		
AT1G79050	DNA repair protein recA homolog 1		445/3				
AT2G36250	Encodes one of two FtsZ proteins,		535/3				
AT3G08530	CHC2 heavy chain subunit of clathrin		662/3				
AT5G58290	26S proteasome AAA-ATPase RPT3		211/3				
AT3G13300	Encodes VCS (VARICOSE)		268/3				
AT3G15730	Encodes phospholipase D alpha 1			1224/3		2675/2	1413/3
AT3G50370	hypothetical protein			747/3		224/2	938/2
AT4G01290	hypothetical protein			699/3			
AT5G37720	atALY-2a			645/3	163/2	321/2	556/3
AT1G18080	WD-40			614/3	295/3	286/2	302/2
AT1G20220	Alba DNA/RNA-binding protein		167/2	584/3		209/2	474/2
AT5G47430	Unknown protein			543/3			845/3
AT1G31817	Ribosomal L18p/L5e			517/3	577/3	217/2	155/2
AT5G48650	NTF2-Like 19			469/3			
AT2G32060	Ribosomal protein L7Ae			440/3	160/2		144/2
AT5G02530	atALY-1a		368/2	435/3	109/2		300/3
AT1G26550	FKBP-like peptidyl-prolyl (EIF3)			426/3		441/2	306/2
AT3G11400				424/3	158/2	431/2	191/3
AT5G15610	Proteasome component (PCI)			423/3	686/2		
AT1G13020	Encodes eIF4B2			396/3		409/2	
AT5G28060	Ribosomal protein S24e			391/3	257/3	217/2	221/2
AT1G76010	Alba DNA/RNA-binding protein			385/3		145/2	
AT2G43030	Ribosomal protein L3 family protein			326/3	455/4	308/2	270/2
AT3G59650	mitochondrial ribosomal protein			321/3	218/2		359/2
AT5G47320	mitochondrial ribosome			319/3	330/2	205/2	
AT1G61870	pentatricopeptide repeat 336			309/3		268/2	205/2
AT3G25150	NTF2-Like 14			266/3	324/3		
AT4G08950	Phosphate-responsive 1 family protein			266/3		382/2	
AT2G44430	DNA-binding bromodomain-containing		174/2	262/3	123/3	234/2	
AT3G13990	unknown protein			249/3			595/2
AT5G35680	Nucleic acid-binding			230/3	140/3		233/3

**Supplementary table 18g. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT5G46160	Ribosomal protein L14p/L23e			178/3	109/3	184/2	196/3
AT3G46560	zinc finger-like protein			175/3			
AT5G48680	Sterile alpha motif			152/3			
AT3G27850	50S ribosomal protein L12-C			131/3	266/2		132/2
AT5G41520	eukaryote-specific protein S10e			225/3	356/3		117/2
AT2G35230	unknown protein			202/3			157/2
AT5G22080	DnaJ-domain superfamily protein			181/3		536/2	351/3
AT4G02840	atSmD1-b			542/2		225/2	1233/2
AT3G08947	ARM			523/2		358/2	
AT4G02400	U3 ribonucleoprotein			497/2		235/2	306/3
AT4G29830	SKI8, VIP3			480/2	112/2	169/2	199/2
AT3G21215	hypothetical protein			478/2			667/2
AT3G07590	atSmD1-a Small			462/2		205/2	786/3
AT5G14320	Ribosomal protein S13/S18 family			407/2	157/4		
AT2G37990	ribosome biogenesis protein			356/2		193/2	229/3
AT1G18630	atGRBP1a			356/2	190/2		161/2
AT3G26400	member of eIF4B			342/2		183/2	
AT4G08580	microfibrillar-associated protein			324/2		192/2	741/3
AT5G38720	ribosomal RNA-processing 7			320/2		136/2	
AT5G66860	Ribosomal protein L25			320/2	392/3	253/2	
AT1G21690	Replication factor			299/2			
AT1G76940	Unkown protein			298/2			934/2
AT1G11480	Unknown protein			291/2			
AT3G03920	H/ACA ribonucleoprotein complex			268/2			
AT4G34620	Encodes ribosomal protein S16			262/2	221/4		
AT1G66260	atALY-2b RNA recognition motif			257/2			192/2
AT5G45600	AtYAF9A,AtYAF9A			256/2			
AT1G07830	ribosomal protein L29 family protein			253/2	274/2	203/2	262/2
AT5G65220	Ribosomal L29 family protein			252/2	252/3	199/2	189/3
AT1G63160	Replication factor			245/2		337/2	
AT2G19385	zinc ion binding protein			242/2		254/2	184/3
AT5G44320	eIF-3			237/2			
AT3G52150	RNA-binding (RRM/RBD/RNP motifs)			233/2	340/4		
AT4G33250	hypothetical protein			230/2	312/3		
AT3G17465	L3 ribosomal protein			230/2		177/2	618/2
AT3G01540	RNA HELICASE DRH1,0			229/2		167/2	602/3
AT2G18040	PPlase			221/2		221/2	213/3
AT2G16060	class 1 nonsymbiotic hemoglobin			207/2		139/2	
AT2G19540	AtCAF1CL6			204/2			157/3
AT1G24240	Ribosomal protein L19 family protein			203/2	277/2		252
AT2G23350	PAB4			193/2			122/3
AT3G23620	Ribosomal RNA processing Brix			190/2			
AT3G47120	GDS1			190/2		253/2	293/2
AT5G22040	unknown protein			174/2			701/3
AT4G14385	EAF6			167/2	310/2		
AT1G63250	DEA(D/H)-box RNA helicase			165/2			223/2
AT2G20360	NAD(P)-binding			162/2		264/2	
AT2G41100	unknown protein			160/2			
AT3G49990	Low temperature viability protein			150/2			
AT5G64260	EXORDIUM			136/2	81/2	207/2	114/2
AT5G55210	hypothetical protein			134/2	245/3	160/2	168/2
AT1G80030	hypothetical protein			134/2			136/3
AT5G50320	ELP3; ELO3		298/2	122/2			386/3
AT4G25550	CFIS2			121/2			317/3
AT1G48650	Unknown protein			119/2			1074/3
AT2G19520	,AtCAF1CL4			119/2			
AT1G52740	H2A variant 3 (H2A.F/Z 3) (HTA9)			98/2	119/2	193/2	
AT4G31120	SKB1			85/2			
AT3G02090	Insulinase					1239/2	501/3
AT3G18035	A linker histone like protein,					806/2	241/3
AT3G54540	member of GCN subfamily,					733/2	358/3
AT4G04020	fibrillin preprotein,					730/2	186/2
AT3G55560	AT-hook protein					702/2	
AT2G44530	Phosphoribosyltransferase					650/2	393
AT1G11910	aspartic proteinase					608/2	367/2
AT3G01280	voltage-dependent		332/2			592/2	395/2

**Supplementary table 18h. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT1G32380	(PRPP) synthase					568/2	
AT4G26110	AtNAPL1		271/2		845/2	541/2	285/3
AT4G01690	PPOX					512/2	285/2
AT4G31880	PO76/PDS5					505/2	
AT2G35390	Phosphoribosyltransferaseprotein					504/2	363/2
AT4G13010	Oxidoreductase					500/2	
AT2G22360	DNAJ heat shock family protein		251/2			493/2	194/3
AT5G10010	myosin-H heavy protein;(source,					491/2	
AT5G04290	SPT5L					471/2	
AT4G22710	member of CYP706A					447/2	815/2
AT5G19760	mitochondrial carrier					419/2	204/2
AT5G21990	36 carboxylate clamp					409/2	556/2
AT1G03860	prohibitin 2					406/2	333/2
AT3G23400	Encodes FIBRILLIN 4 (FIB4)					395/2	242/2
AT3G01340	Transducin/WD40					391/2	
AT2G41040	S-adenosyl-L-methionine-dependent		262/2			390/2	319/3
AT2G42130	Plastid-lipid					385/2	
AT2G30050	Sec13					375/2	
AT4G26630	chromatin-associated protein t					374/2	
AT1G50840	DNA Polymerase gamma2					365/2	
AT1G44170	ABA and dehydration					360/2	104/2
AT3G22370	AOX1a					360/2	398/2
AT1G22530	PATLs					350/2	
AT4G28510	prohibitin 1 (Atpb1),					348/2	316/2
AT5G50850	Transketolase		462/2		102/2	347/2	264/2
AT4G12080	AT-hook				104/2	346/2	158/2
AT3G62360	Carbohydrate-binding-like					346/2	
AT5G56950	AtNAPL3,				553/4	345/2	
AT5G13450	Mt ATP synthase					340/2	
AT4G17530	AtRabD2c encodes a Rab GTPase					337/2	
AT5G51590	AT hook motif DNA-binding		182/2			337/2	132/2
AT5G08420	RNA-binding KH domain-containing					333/2	171/2
AT2G20530	prohibitin 6					333/2	
AT3G46830	RAB GTPase homolog A2C					329/2	213/2
AT5G35910	RRP6,					329/2	
AT4G17720	RNA-binding (RRM/RBD/RNP motifs)					321/2	
AT1G09630	Encodes a putative GTP-binding					314/2	
AT4G39960	Hsp4DnaJ family protein					314/2	169/2
AT4G04910	N-ethylmaleimide sensitive factor,					312/2	
AT5G53560	AtCBR					312/2	335/2
AT3G45830	B-binding-like protein					310/2	342/2
AT3G45830	B-binding-like protein					310/2	342/2
AT3G13290	varicose-like protein		253/2			306/2	356/3
AT2G38550	Transmembrane proteins 14C					300/2	268/2
AT4G01395	conserved oligomeric Golgi complex					298/2	
AT1G76810	(eIF-2) family protein					296/2	
AT3G12130	KH domain-containing protein					295/2	
AT4G15570	Similar to yeast Sen1					294/2	
AT5G61790	calnexin 1					292/2	345/2
AT5G13430	Ubiquinol-cytochrome C reductase					290/2	
AT1G74020	Encodes AtSS-2 strictosidine synthase					283/2	
AT5G66420	TIM-barrel signal transduction protein					279/2	
AT2G30920	complement Saccharomyces c. coq3					277/2	
AT4G35000	APX3					275/2	306/2
AT1G04710	EC2.3.1.16 thiolase					273/2	
AT1G14900	subgroup of HMGA					266/2	123/2
AT5G22010	Replication factor					264/2	
AT3G22320	NRP(A/B/C/D)5					262/2	
AT1G48620	histone H1/H5 family member				534/2	258/2	
AT2G44610	Encodes a GTP-binding protein					256/2	
AT5G14220	PPO2					255/2	255/2
AT1G49480	DNA-binding protein					255/2	
AT1G48610	AT hook motif-containing protein					254/2	
AT1G28200	VirF-interacting protein FIP1,					251/2	
AT3G26080	plastid-lipid associated protein PAP					251/2	
AT3G45970	EXPANSIN-LIKE					250/2	
AT1G51560	Pyridoxamine 5-phosphate oxidase					246/2	172/2
AT2G17190	ubiquitin receptor protei					244/2	980/2



**Supplementary table 18i. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT5G63400	adenylate kinase.,					241/2	
AT4G36020	cold shock domain protein					239/2	
AT5G48810	cytochrome b5 isoform					238/2	
AT5G08740	NADPH dehydrogenase					234/2	
AT4G05410	WD40-repeats					231/2	108/2
AT3G59820	LETM1-like protein;(source,					231/2	
AT4G33510	mRNA is cell-to-cell mobile					231/2	197/3
AT3G10650	Nup136					231/2	
AT3G15190	chloroplast 30S ribosomal protein S20				470/2	218/2	192/2
AT1G19870	microtubule-associated					217/2	
AT5G12470	UvrABC system C protein					215/2	175/2
AT4G19006	Proteasome component (PCI)					215/2	
AT1G76850	exocyst complex gene family					214/2	
AT4G32470	Cytochrome bd ubiquinol oxidase					213/2	
AT1G67360	small rubber particle protein homolog					211/2	
AT3G48680	gamma carbonic anhydrase-like					210/2	
AT3G25920	plastid ribosomal protein CL15,				408/4	209/2	214/2
AT3G25920	plastid ribosomal protein CL15,				408/4	209/2	214/2
AT3G22520	spindle assembly abnormal protein					209/2	
AT3G21140	Pyridoxamine 5-phosphate oxidase					208/2	
AT2G01470	Sec12p-like					206/2	
AT1G47260	gamma carbonic anhydrase					206/2	
AT5G63510	gamma carbonic anhydrase-like protein					204/2	
AT5G13530	RING E3 ligase i					203/2	382/2
AT5G55660	DEK					202/2	
AT2G27020	20S proteasome alpha 7 subunit PAG1					202/2	
AT2G22250	plastidic aspartate aminotransferase					202/2	
AT5G25450	Cytochrome bd ubiquinol oxidase					201/2	
AT2G42610	LIGHT-DEPENDENT					199/2	245/2
AT4G38970	tyrosine-phosphorylated					199/2	
AT4G22240	Plastid-lipid associated protein PAP					198/2	
AT2G20940	transmembrane protein					197/2	
AT3G51800	DNA-binding protein G2p (AtG2) mRNA					196/2	
AT3G21400	dynein beta chain					196/2	
AT2G35490	Plastid-lipid associated protein PAP					195/2	
AT4G29130	Encodes a hexokinase (HXK1)					195/2	
AT5G23060	chloroplast-localized protein					194/2	165/2
AT5G05000	GTPase protein					194/2	
AT3G55005	unknown function					193/2	170/3
AT5G13200	hormone mediated regulation					192/2	
AT2G30490	cinnamate-4-hydroxylase					191/2	184/3
AT5G42150	Glutathione S-transferase					190/2	104/2
AT5G59840	GTP-binding family protein					186/2	
AT3G63410	MPBQ/MSBQ methyltransferase					185/2	
AT1G47420	5 of mitochondrial complex II					184/2	
AT1G59990	DEA(D/H)-box RNA helicase					184/2	181/2
AT1G67230	nuclear coiled-coil protein					182/2	188/2
AT1G31780	oligomeric golgi complex subunit					181/2	
AT2G22780	NAD-malate dehydrogenase					179/2	
AT4G35860	GTP-binding protein ATGB2,					179/2	
AT3G26070	Plastid-lipid associated protein PAP					179/2	
AT5G42960	outer envelope pore 24B-like protein					178/2	206/2
AT2G22230	Thioesterase superfamily protein					174/2	
AT5G66510	mitochondrial gamma carbonic anhydrase					173/2	
AT3G19820	24-methylenecholesterol					172/2	
AT3G49100	SRP9/SRP14					172/2	
AT1G07090	LIGHT-DEPENDENT					171/2	
AT1G32990	Pale green plants; Reduced growth rate				345/4	171/2	132/2
AT3G46220	E3 UFM1-protein ligase-like					169/2	
AT5G53310	myosin heavy chain-like protein					168/2	
AT4G23650	calcium dependent protein kinase 3					168/2	
AT4G05400	copper ion binding protein				479/3	167/2	
AT2G43640	SRP9/SRP14					167/2	
AT2G39670	Radical SAM superfamily protein					167/2	
AT3G01800	Ribosome recycling factor					166/2	
AT4G27690	vacuolar protein sorting 26B					165/2	
AT1G14850	Nup155,					164/2	341/2
AT4G17890	ARF GAP domain (AGD)					161/2	

**Supplementary table 18j. List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS.** The numbers indicate the respective average MASCOT score and how many times the proteins were detected in three independent APs. The proteins that were detected in less than two out of three AP are not listed.

AGI	Description	GS-MRG2	MRG2-SG	GS-MRG1	MRG1-SG	GS-MRG2 w/o benzonase	PTB2-SG
AT4G17890	ARF GAP domain (AGD)					161/2	
AT3G22950	ARF GTPase family					160/2	
AT4G31160	DCAF/DWD					160/2	213/2
AT5G08540	RNA small subunit methyltransferaseJ					159/2	
AT5G60730	GET3c					157/2	170/2
AT3G11730	Rab GTPase family of proteins					155/2	
AT1G02930	GSTs					154/2	
AT5G24650	HP3Tric1					154/2	
AT4G33350	Tic22-like					151/2	
AT5G53530	VPS26					151/2	
AT4G27500	interacts with H <sup>+</sup> -ATPase					148/2	
AT5G53400	BOBBER1 (BOB1)					146/2	
AT1G33780	putative (DUF179)					146/2	
AT4G12790	triphosphate hydrolases superfamily		118/2			145/2	128/2
AT2G35780	serine carboxypeptidase-like 26					143/2	
AT3G18820	RAB GTPase homolog G3F					143/2	
AT4G01660	ABC1-like protein					141/2	
AT1G19140	ubiquinone biosynthesis COQ9-like					139/2	
AT2G40430	homolog of yeast NOP53					138/2	
AT5G65270	RAB GTPase homolog A4A					138/2	
AT1G07320	encodes a plastid ribosomal protein L4				416/4	136/2	112/2
AT2G45140	VAMP-Associated protein					136/2	
AT3G06050	mitochondrial matrix peroxiredoxin					135/2	198/2
AT1G06530	MITOCHONDRIAL DIVISION FACTOR2		164/2			134/2	176/2
AT1G60780	Clathrin adaptor complexes					133/2	
AT1G15440	ribosome biogenesis co-factor					130/2	
AT1G49040	WD-40 repeats					130/2	189/3
AT1G64550	GCN subfamily					129/2	
AT5G26710	Glutamyl/glutaminyl-tRNA synthetase,					129/2	
AT5G18660	3,8-divinyl protochlorophyllide					128/2	
AT2G17840	drought-inducible gene					127/2	
AT2G14740	vacuolar sorting receptor					126/2	
AT4G18040	eIF4E protein.					125/2	217/2
AT4G28440	OB-fold-like protein					122/2	
AT3G52850	Vacuolar Sorting Receptor-1 (VSR-1)					122/2	
AT1G11580	methylesterase PCR A					122/2	
AT2G16485	NERD					116/2	842/3
AT1G13220	Encodes a nuclear coiled-coil protein					116/2	
AT2G34970	Trimeric LpxA-like enzyme					115/2	
AT1G16890	UBC36/UBC13B					115/2	
AT5G11560	catalytics					112,/2	
AT3G44330	M28 Zn-peptidase nicastrin					100/2	

**Supplementary Table 19** Alternative splice forms were coamplified and quantified using a Bioanalyzer. Ratios were normalized to the corresponding control Col-0. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS

## At3g58510

Col-0	<i>lws1</i>	<i>mrg1/mrg2</i>	<i>mrg1/mrg2/lws1</i>
1	2.00	0.14	1.88
1	1.79	0.82	1.75
1	1.50	0.47	2.12

## At1g09140: SR30

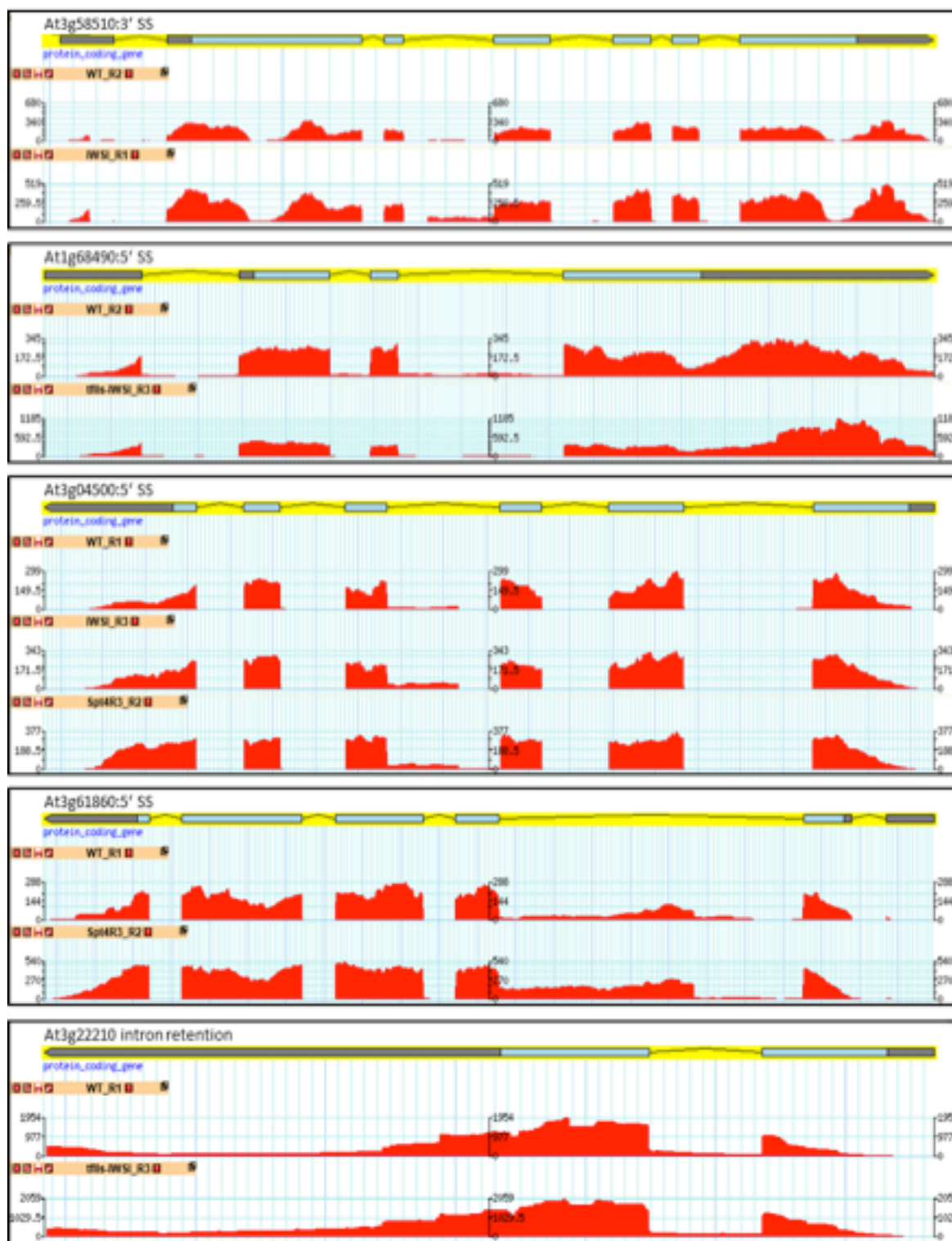
Col-0	<i>lws1</i>	<i>mrg1/mrg2</i>	<i>mrg1/mrg2/lws1</i>
1	1.97	0.62	1.47
1	3.94	0.88	3.76
1	2.16	0.68	1.79
1	3.24	1.29	3.10

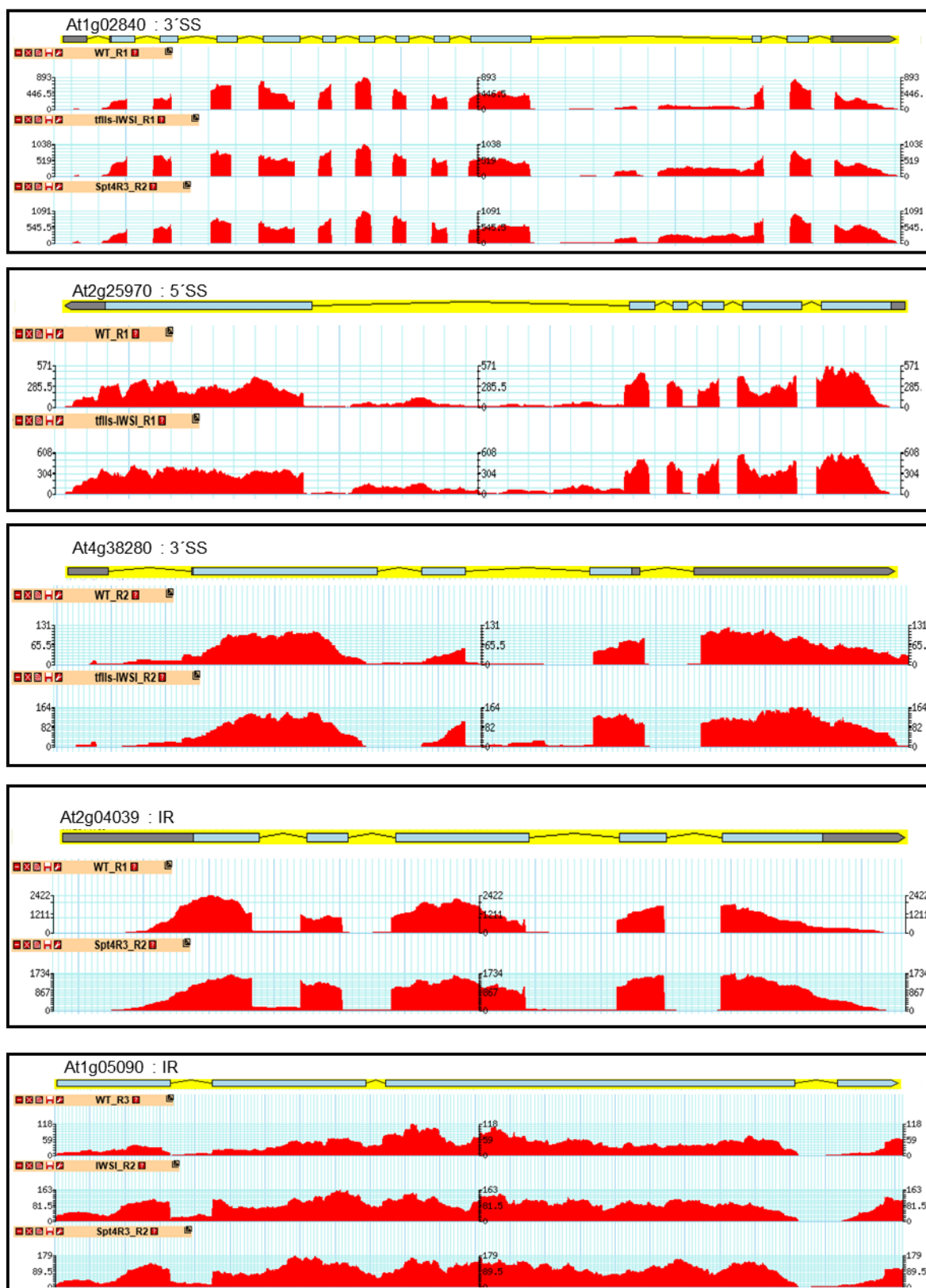
## At1g68490

Col-0	<i>lws1</i>	<i>mrg1/mrg2</i>	<i>mrg1/mrg2/lws1</i>
1	0.23	1.24	0.26
1	0.41	1.34	0.30
1	0.60	2.75	

## At3g58510

Col-0	<i>sdg7.2</i>	<i>sdg7.3</i>	<i>sdg8</i>
1	0.64	0.8	1
1	0.754717	0.830189	1.09434
1	0.677966	0.508475	0.847458





**Supplementary figure 1. RNAseq cover reads of differentially alternatively spliced genes in Col-0 and the mutant(s) that show altered AS events.** Gene models depicting the detected differentially spliced genes: At1g02840, At2g25970, At4g38280, At2g04039 and At1g05090 and corresponding plotted RNAseq reads in Col-0 and the mutant(s) that show AS events differentially affected. Coverage plot reads within the AS regions are shown in red. A significant difference in 3'SS event for At1g02840 (tenth intron), 5'SS for At2g25970 (fifth intron), 3'SS At4g038280 (third intron), intron retention At2g04039 (first intron) and intron retention for At1g05090 (first intron) between Col-0 and *tef* mutants can be detected.



**At3go4500: alt 3' SS****Major splicing isoform sequence:**

AAGACTCTTATTTTTCTGGTACGAGTTTGACCACAAAACCTAACACTGAATACTTATCTTTGTGATTTCGAGTAAGAATGA  
 TCCGATGATCTAAGGTCAACCATGCTCTGTTCTGAACATACTATGTGTTTTGGTTTGCAAATGCATCACAGTTGACT  
 TGGCCATATTGAAGGTGGGGACAAATATAT

**Minor splicing isoform sequence:**

TCCCCACCTTCAATATGGCCAAGGTCATTAGAGATAAGCGGACTGGTAAAACCAAGG

**At3g22210: Intron retention****Major splicing isoform sequence:**

TGCCGTTAGCTTTGCCGAGATCAGAGAGGTTCACTGTTATGAACTACTCCTTATTCTCAACATTTTTGATCTTTTACA  
 TATAATTTTAAGAAAAAATCAAGTGTGTGTGTGCAGAAATCGTTTAAAGATGTGTACGACTCGGA

**Minor splicing isoform sequence:**

TTCTTGAATCGTCGCAGTCTCGGTCGTAACATTCTATGCCGTTAGCTTTGCCGAGATCAGAGAGAAATCGTTTAAA  
 GATGTGTACGACTCGGA

**At1g02840: alt 3' SS****Major splicing isoform sequence:**

TACGACTAGAAATGGAAATTAATATCTGGATGGAAATTGATTGTTGACAAGTGTTCGTCTAAGAGATGGATAACAGT  
 GGAGACATTGGATCACTTGGATCACAATATTATATCGGGATTTCTGTAAAATATATTGGCTCGATGGATTGACAATA  
 TGGAAATCTGGGTTCTCTTGGGACGTCCGTGGCTCATTGGCAACACAAGTTTTTTTTGGCCACATGGCTTATAAAACC  
 TCTGTCCTATCACCTATGTTTTAACTAAGTAGCAGAATAGTTTGGTTTATGTTTCTTTTTTTTTATTTGTTGCAACTTCT  
 TAATCTCTGTGAGATAGAAGGAGAGGCTCCAGGACCTTGTGAACAGTATAAAACACAACATGTTTGGATTTTTGAAT  
 CTGAGTTTTCTTTCTTGGACTTTTGCAGATCGCGTTCAAGATCAAGATCTCCTC

**Minor splicing isoform sequence:**

TCGAAGTCAAGGTCACCGTCTCCAAGAAGATCGCGTTCAAGATCAAGATCTCCTCTACCTTCAATCCCGCGGCCATG  
 GCGGCCGGGAGCATGCGACGTGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAGTGGCCGTCGTTTTTA  
 CAACGTCGTG

**At2g25970: alt 5' SS****Major splicing isoform sequence:**

AATCATCAGTGGCGAGGTATGTTTGATTGTTGTGTTTGTATAGATTATGACAGGCTTCTGCTGTGTCTATTTTTATA  
 CATGAATTTGATATTGTGTTGTGATTACATTATTTAGTTGTCATTAGCTTTAGGTCTATCTAGTATACGTCATACTCTAT  
 TATTCAGATAATAGGTGATGTAGGTTATACTAGTTTTCACTTAGTCAATATATTAGCTGCTAAAACCTATACTCTTCTT  
 ATGCCCTTCATTTCTTTGATATTTGCTCCACGATAATTGCATCTGCATTGGATTGCATTGTTCTTCTTCCGCTTATT  
 TGCAGGGGAAGGTTCTTTATTTTATTTTAAACCACGTGAAGATATTTTTGTGATTTTCTTAACCTACAAAAAATATTCC  
 ATCTTGATATATCTTACTTTGTTATTGCATTGTAAGATTATATGTTGCTTCTTCTTCTCTGCTCATGCTCATTTTGTG  
 CCGCTACCTTCTACACGAGTTTCATGTTTCTGTTACTATCTTATTGTGGCAGTTTTTATGGGCATTTCGCTTATATCTA

CTTTTCATCTTTTTCAATATCCAATTTACAGTGTGCTTAGCATGATTCTTCATAACTCCCATTGCTACCAAGAAATATTC  
 TTTAGCATATTCACCTTTTCAAATGCTATTTTCATAGTAGAAATCTTGTTGTCTATTTTTGGAGTATATATGCAGACATTT  
 TCATCGCTTCTGGTACAGCCAAGGTGGCCAATCCTACATTGGTTTGGGATGCCCTTTGTGAGGCTGCTGACTAGTTT  
 CAGTCCAACCTGCTGTGATAGTGTCTAAACTGTTTTATACAACTGTTTTT  
 ATACATACTGTTTTTGTACATTGTGTTTCATCATTCTTGGGTCAACATAACTTCCGTTTCAGCTAGTACTTCTAGTTTCA  
 ATCATCAAGGTACCACCAAAGGTGGGCGACTGACATTTGAAGACCTATTAGCCGTTGTGTTGTGTCAATAATTTGCT  
 AACCGAGTAACTGTTATTATTCTGCCTTGGTACTTTCCCAATTTTTCTTAAATCCGTTTTTCTTTACCTTCACCATTAT  
 ATTCTGGTAATATACTCCATGAGTATTGACATTGTTGAACACTAGACTCTTATAGTTGGTATTAGGTTATTTTGTGCA  
 CAACCCTTTGAGAAAATAGGTTCCATGTGTATATATATGTTGCTTTTTGTCTGCAAAAAGCTTCTATAGTATTGTTGTT  
 TTTCTTGCACCCATTTATCTCGAATAGGGAAAGGGATGGGCTTCTGCAGGAGTAACCCAATAATTT

**Minor splicing isoform sequence:**

ATCAGTGGCGAGAACCGTATGAGAAAACCTCAGCAATGGGTGGA

**At1g19400: Intron Retention****Major splicing isoform sequence:**

GAGTACGGGAATTACAAAGAGGAATACAGAAGGAGACATGTATTGGAGAAGTGAGTGAAT  
 TCATCATTCTAGGAAGACACATCATCATGCCTTTGGTCTGCAGGCAACAACAGAGTACGGGAATACAAAGAG  
 TATTCTTCGGAGATAGACCATTTTAAATT  
 CATATGCGTGATAAAACCCTCTTTCTTCGCAATGCTCACAGTTTTCTTGAATATTTTTGTTTGTGTTCAACACATTAATTA  
 CTATCCCTGATGAAAAACTTATAAGTCATAAAATGTTGTGATGGTTTTTGGATCCTCTTCTTTCCCATTTGAATGAATA  
 CATGTTTGTGGTAGTTTTATGAAGATAAAAGTCCTTAATAAATTGAAGTGGACCAGTCTCCTCCTTATAAATTTGGCA  
 TGTGGCCTAGACACAGTTTACGAATGAAGAGATAAATAGGAATCTCTCATTGCTCCTTTATCTTTAAGCTCTTATC  
 CTTTTGTCTGCTGCTTCTATGACCCACATATGTTGGTCACTTACAATCCCCAATATTTGCTTTGGTGAATGTT

**Minor splicing isoform sequence:**

AGAAAGACACCTCCTATGCACTTGATCTTCAGGCAACAACAGAGTACGGGAATTACAAAGAGGAATACGAAGGAGA  
 CATGTATTGGAGAAGTGAGTG

**At4g38280: alt 3'SS****Major splicing isoform sequence:**

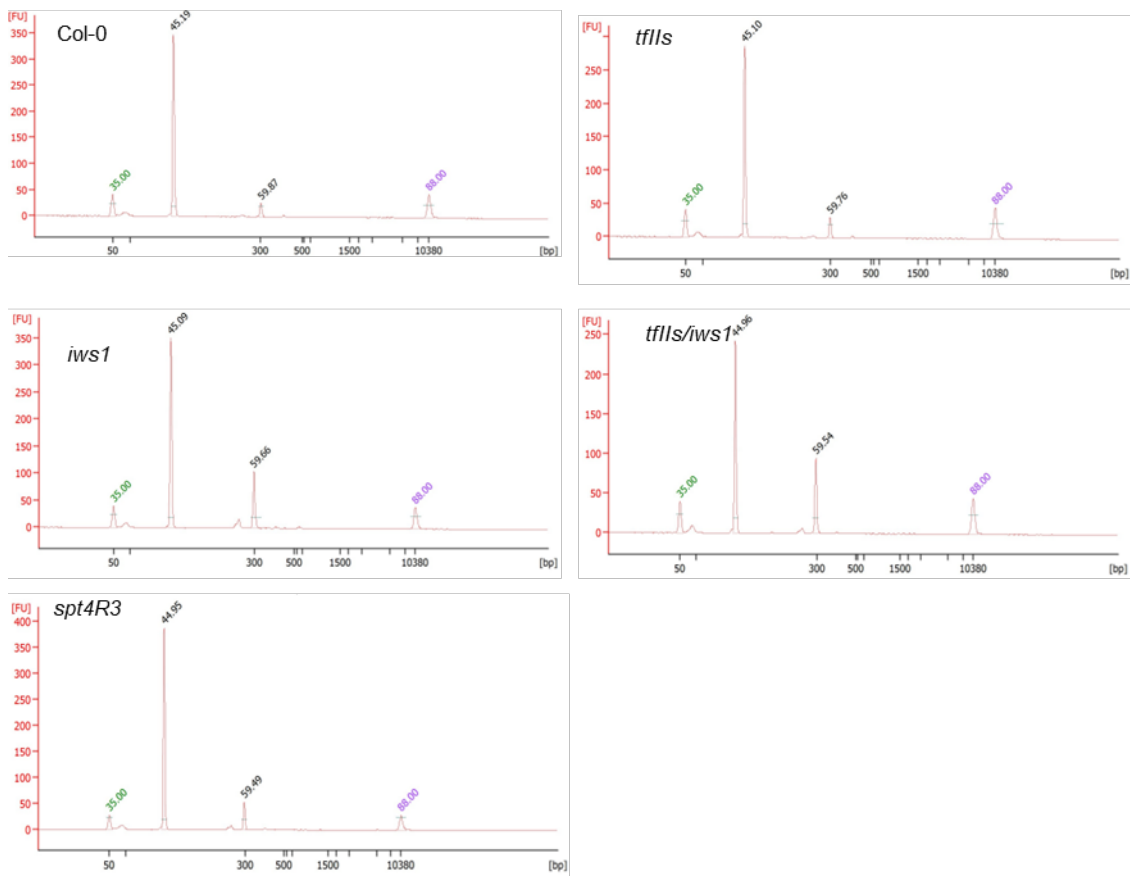
GTTCTCTCTCGAGGAGCTACTAGTACTGCTTCTACGTACCACATCTTTCAAATCCTACCATTGCTGAATCATATCTGT  
 TTGGCACAGTTCTAGTTGACCATCCATAAGTGGATTATTCATGCTGCTGCTTAAGCCATTTTTTCCATATCTGCTGAC  
 AACTATATAGTTTTAGTCACATGCTTTTAGGTTTGTAGCTATGAACCATCTGGAATCAAGTACTTGCTAGAAATAATG  
 TTGTTGATCCTCTGTTGCAGCGAGAGAGATGCAGCGCGTAGCAGCTTTAAA

**Minor splicing isoform sequence:**

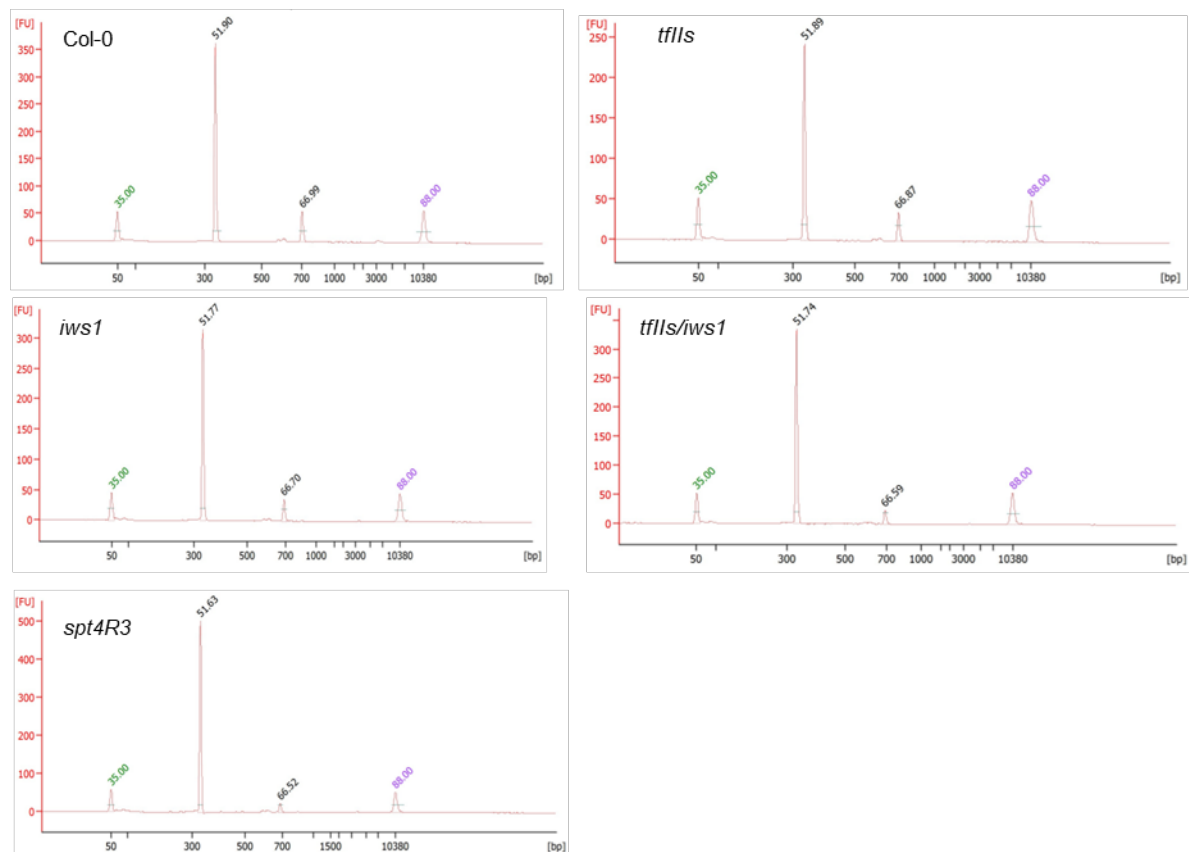
TCTCTCTCGAGGAGCGAGAGAGATGCAGCGCGTAGCAGCTTTAAAAGGCTGCTGCTTAAGCCATTTTTTCCATATC  
 TGCTGACAACTATAGAGCGAGAGAGATGCAGCGCGTAGCAGCTTTAAA



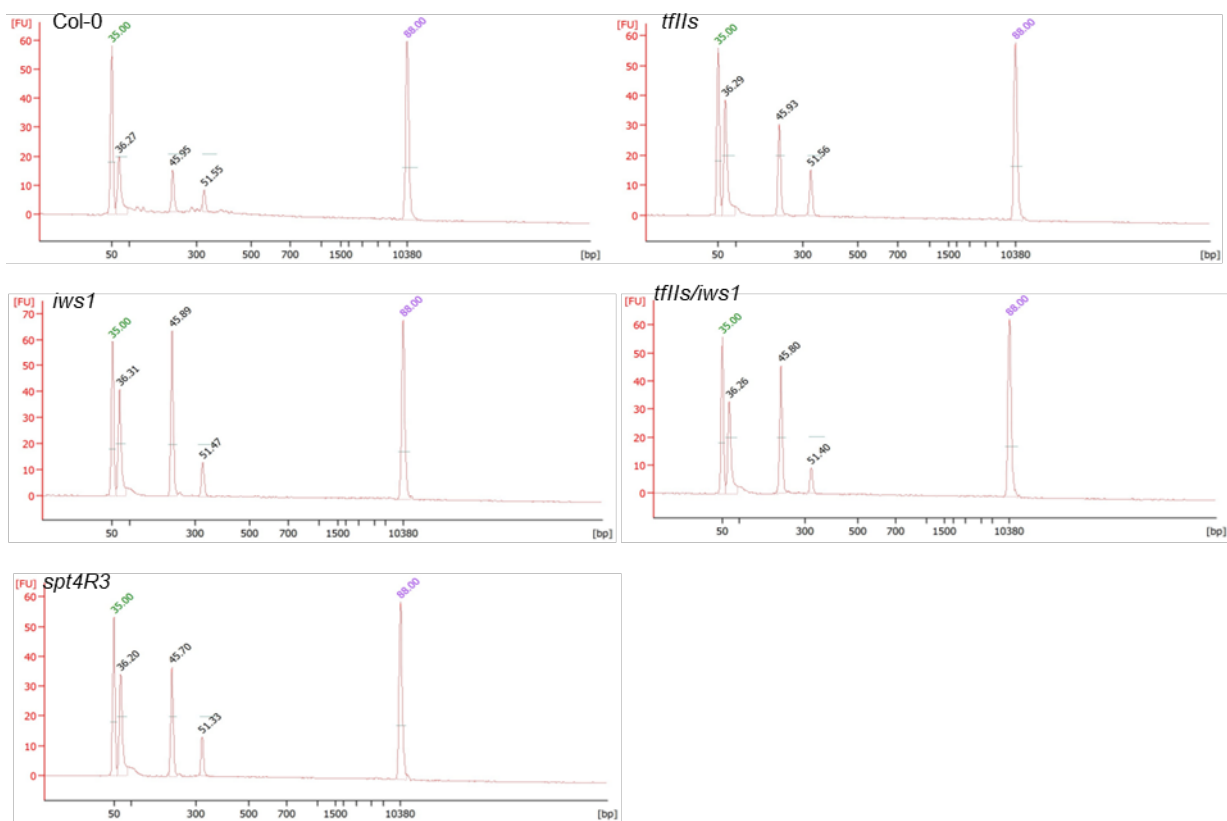
A



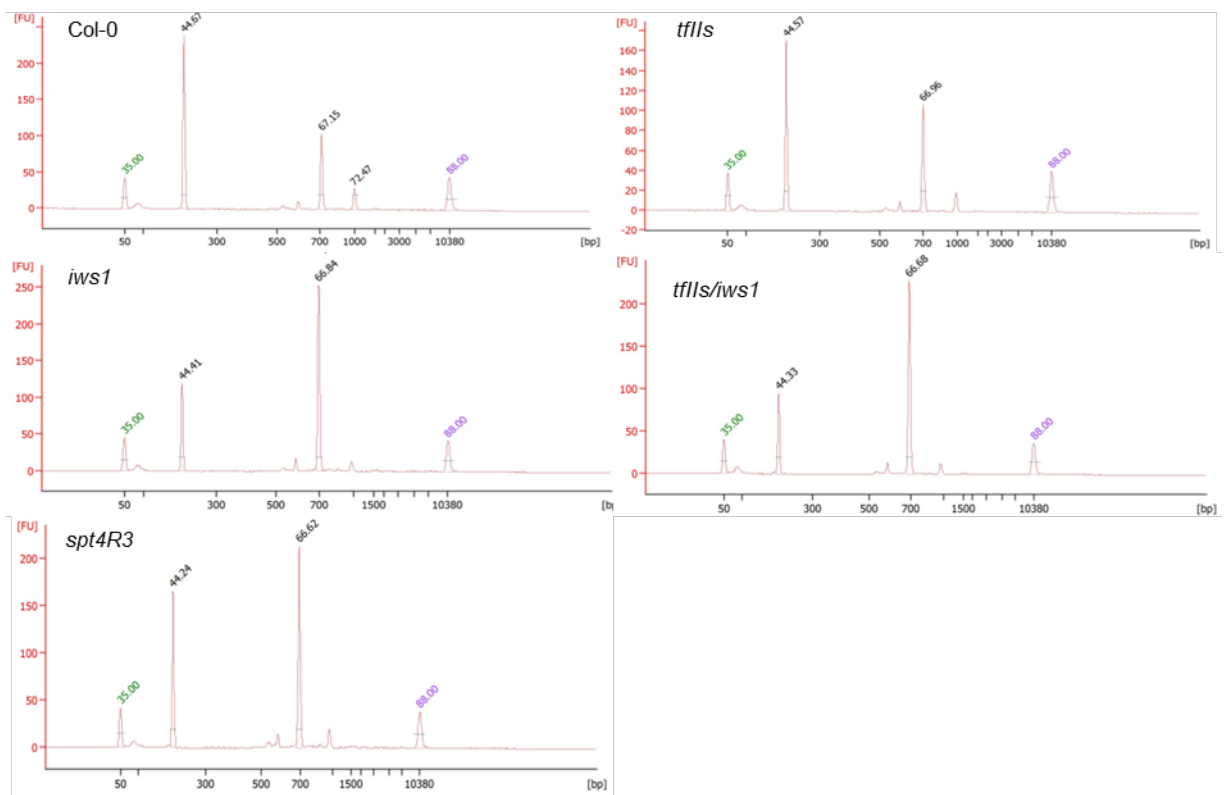
B



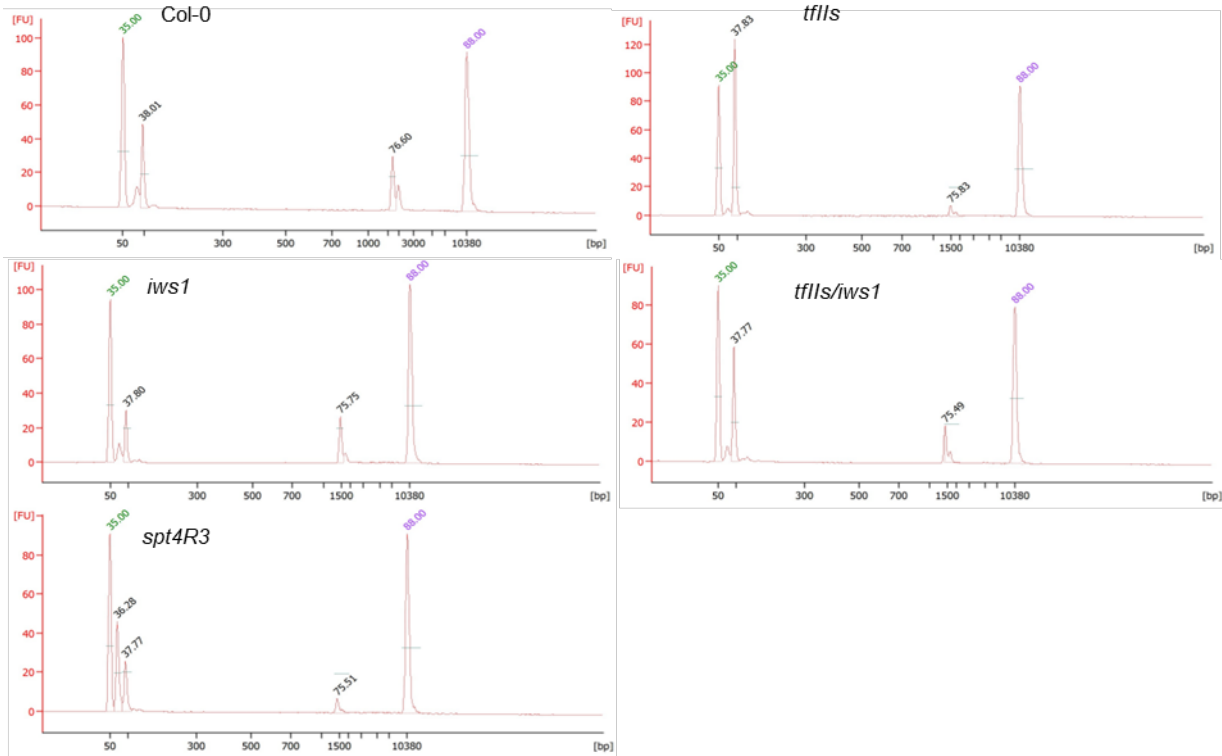
C



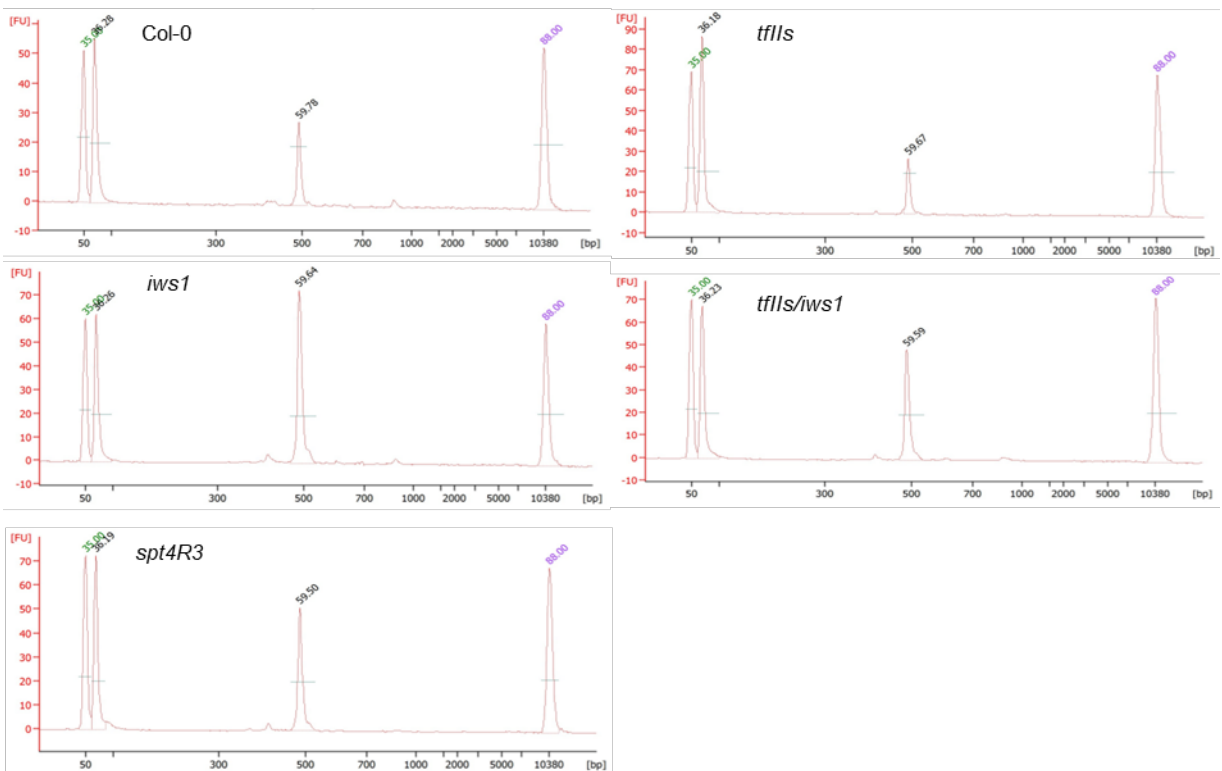
D



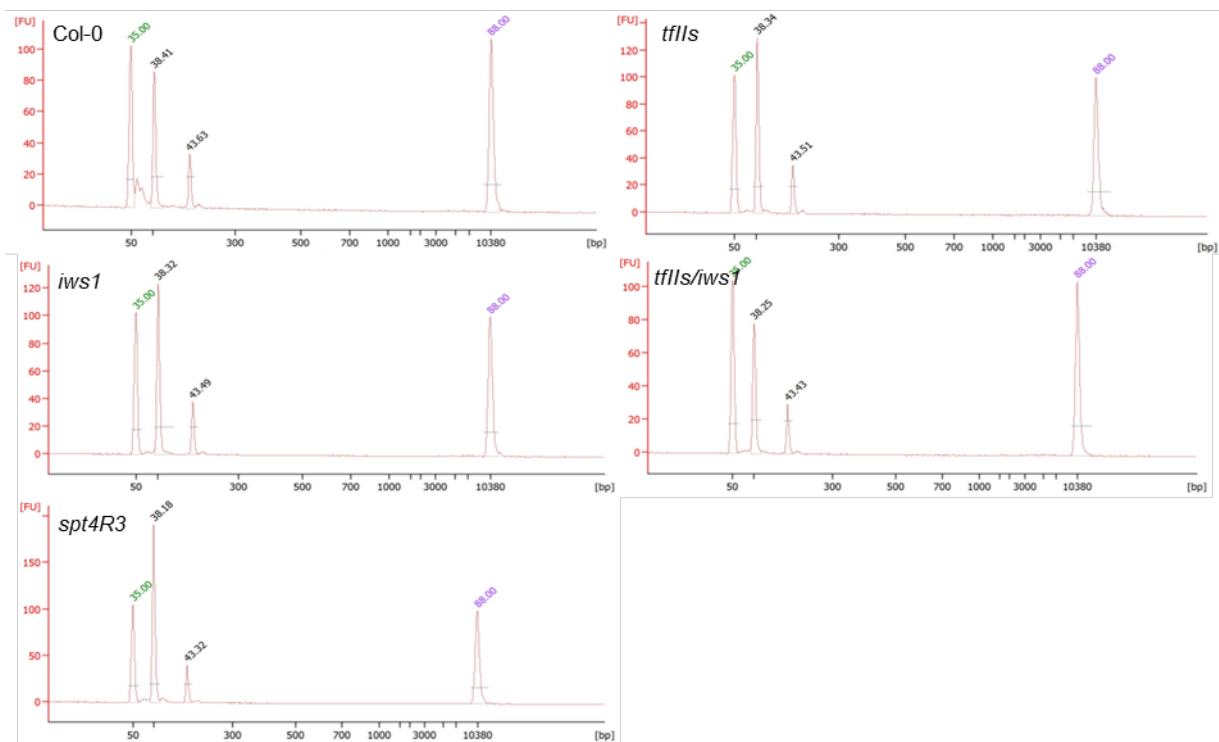
E



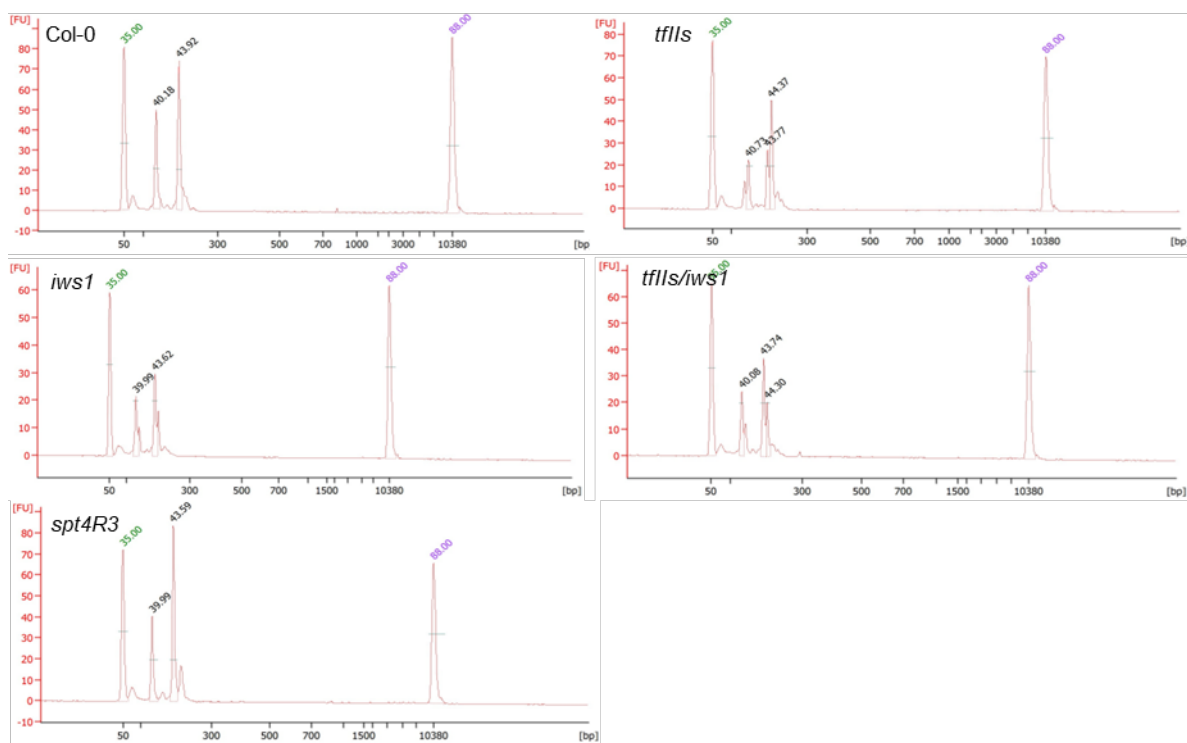
F

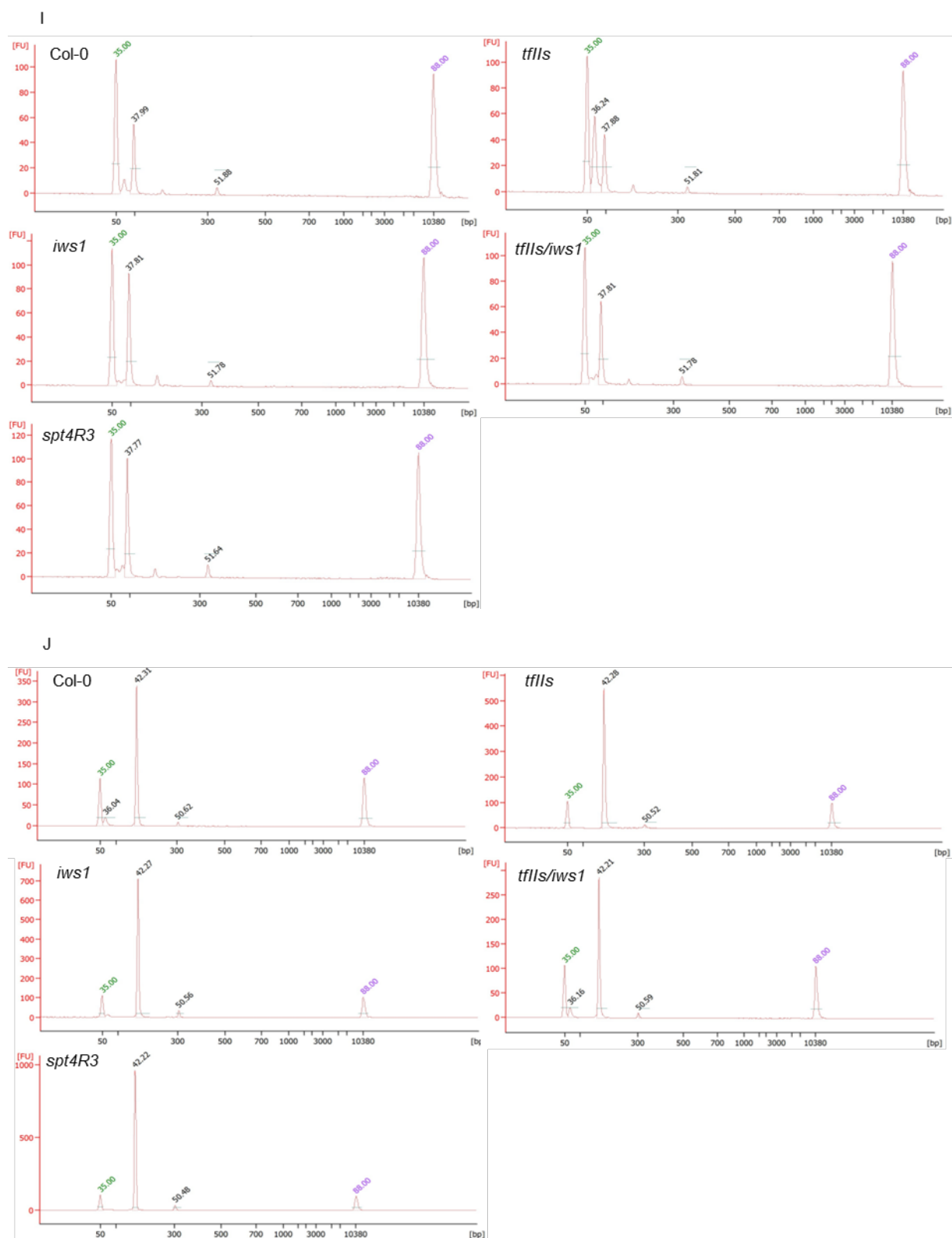


G

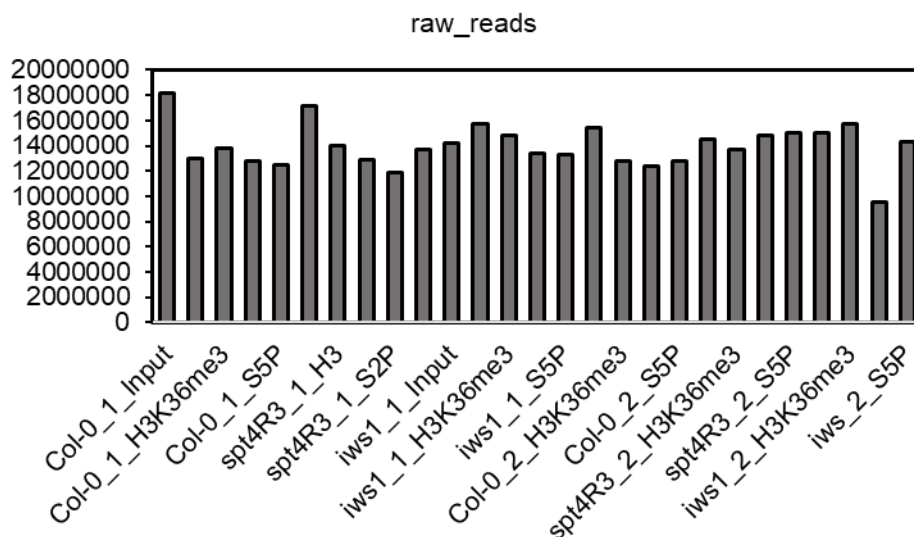


H

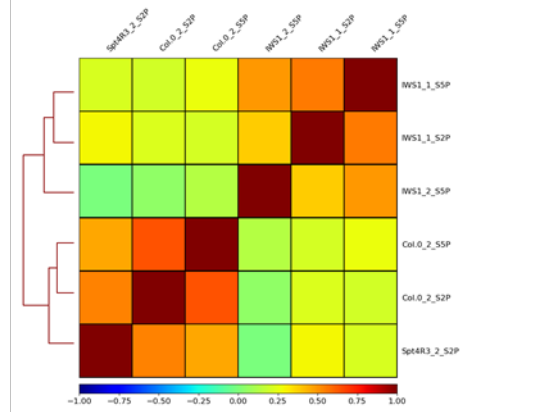
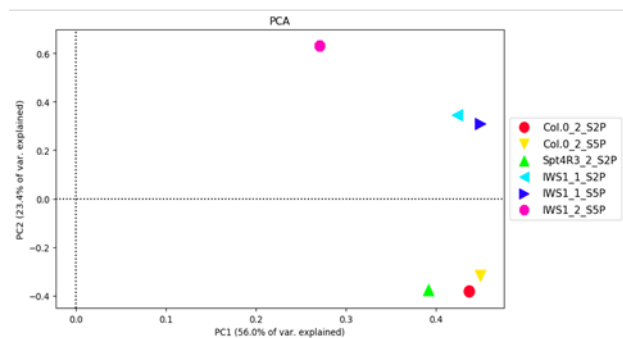
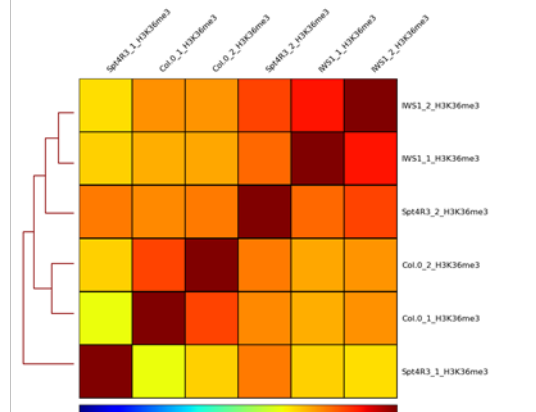
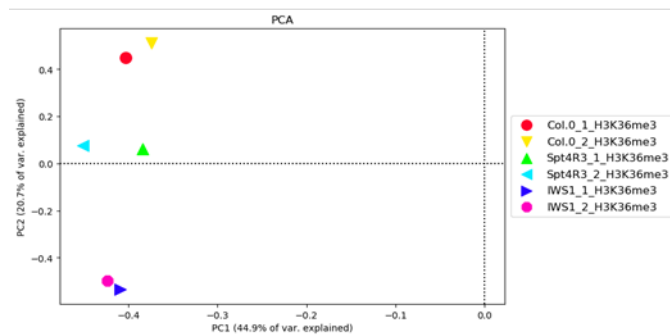
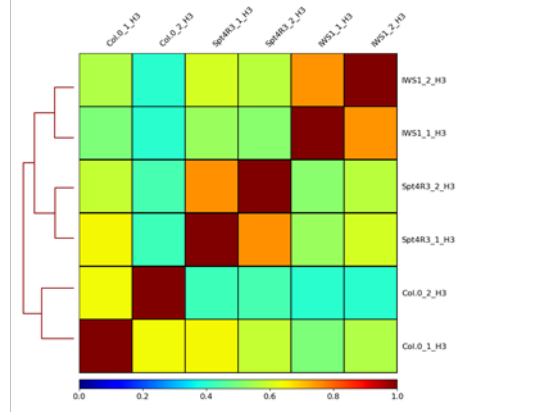
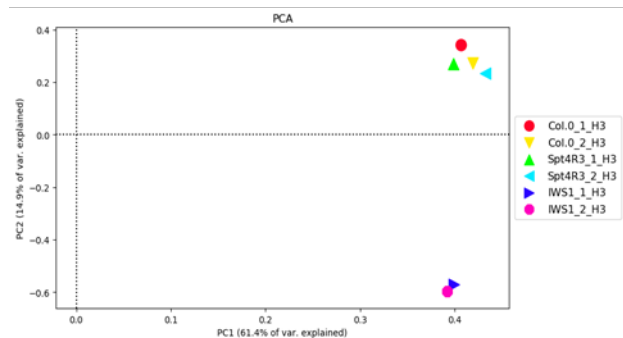




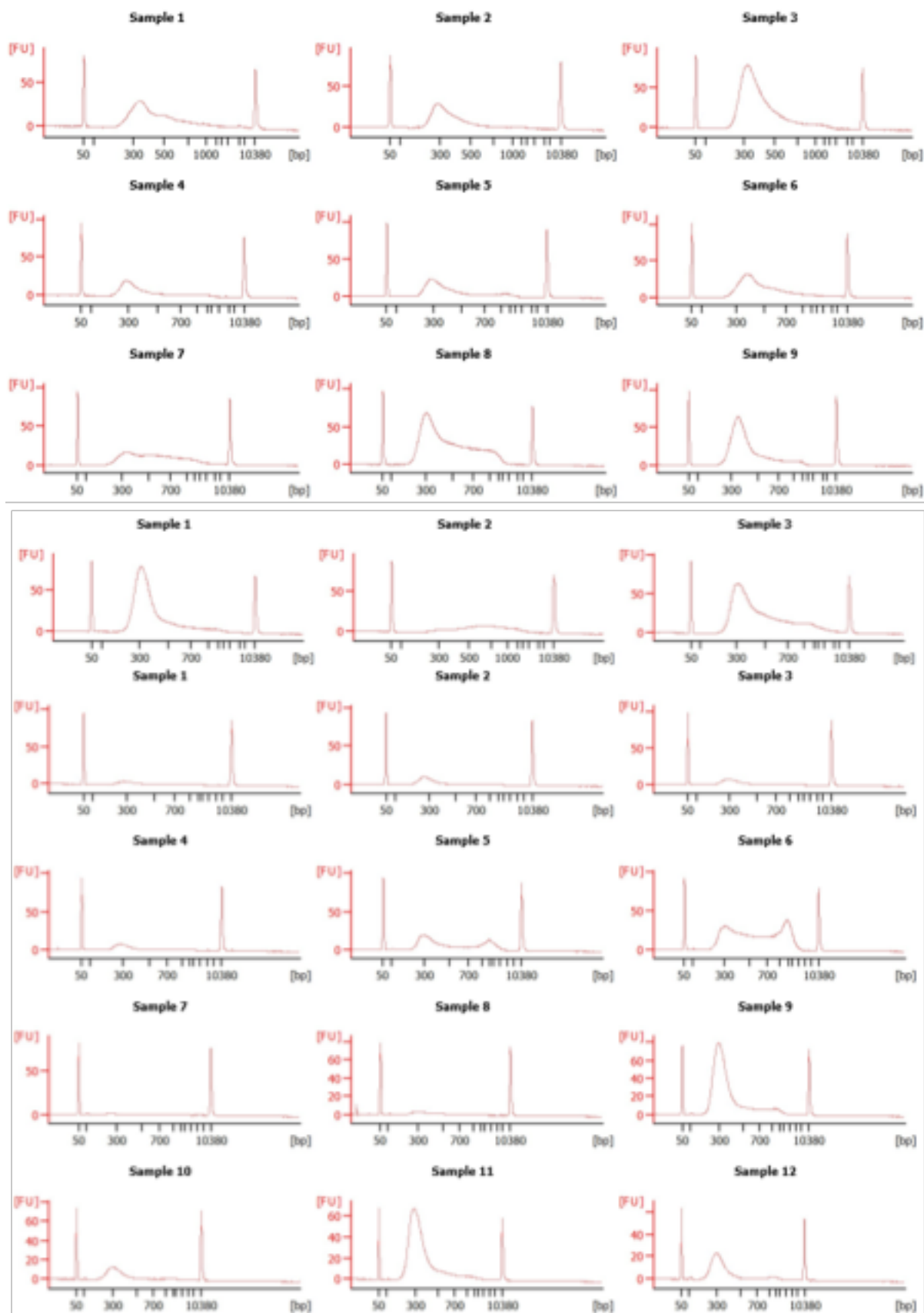
**Supplementary Figure 3.** Quantitative analysis of the ratio of different splice variants in At3g58510 (A), At1g68490 (B) At3g04500 (C), At3g61860 (RS31) (D), At2g25970 (E), At1g02840 (SR34) (F), At3g22210 (G), At1g50902 (H), At4g38280 (I), At2g04039 (J). Using RNA isolated of the different genotypes (Col-0, *tflls*, *iws1*, *tflls/iws1*, *spt4R3*). DNA fragments were amplified by RT-PCR with gene-specific primers spanning the transcripts. The obtained DNA fragments were analysed using a Bioanalyser. Numbers on top of the peaks give the retention time. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS. Ratios were normalized to the corresponding control Col-0. The respective nmol/l concentration, used to calculate the splicing ratio, for every sample of every replicate are in the Supplementary table 10.



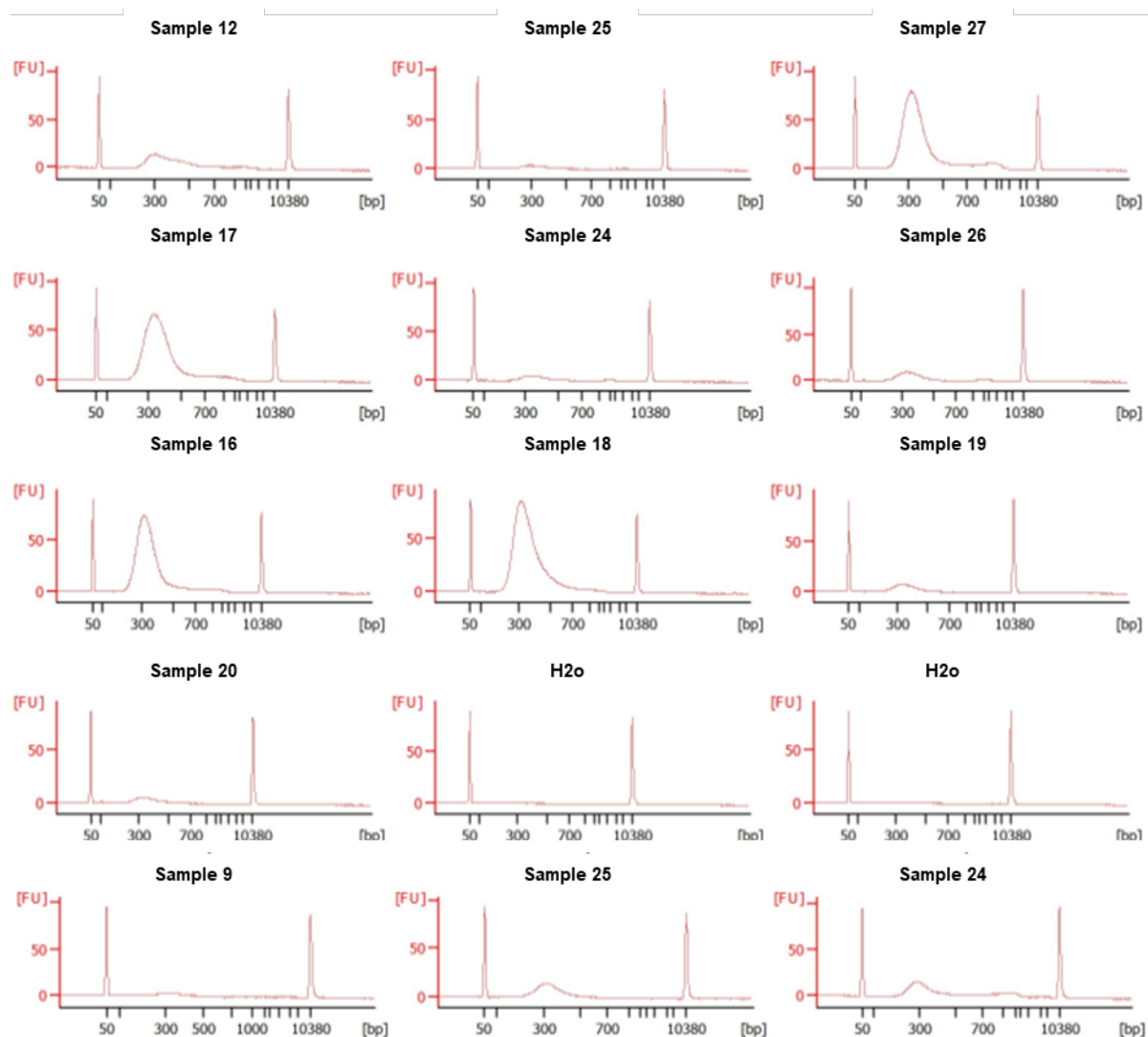
**Supplementary figure 4.** Number of reads generated by Illumina Sequencing in the analysed sample Col-0, *spt4R3*, *iws1* with H3K36me3, S2P and S5P antibody (Mio. reads per sample). Two biological replicates were sequenced for every genotype.



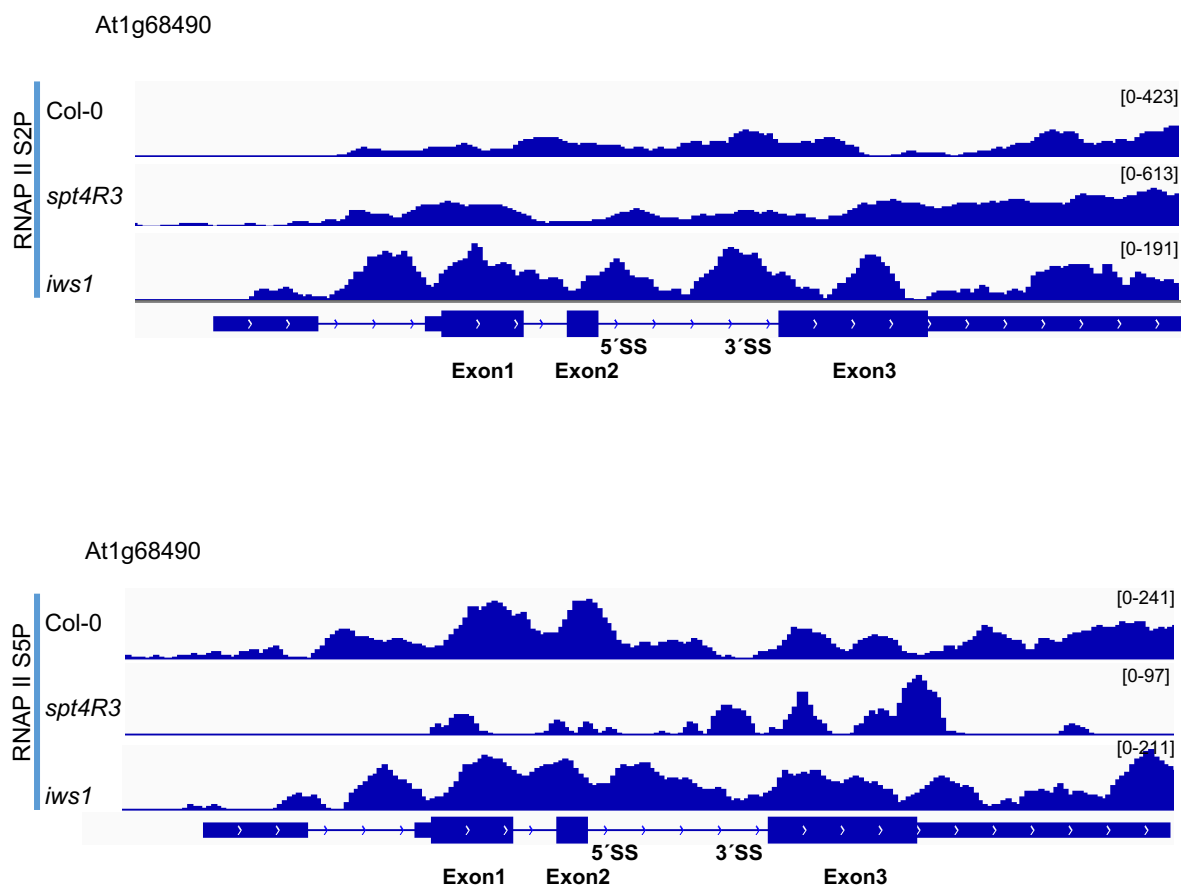
**Supplementary Figure 5. Principal Component Analysis (PCA) and Spearman correlation of datasets obtained by ChIP-seq for individual biological replicates.** The distribution pattern of H3, H3K36me3, S2P and S5P, is reproducible across individual samples representing biological replicates ( $n = 2$ ). The analysis of individual biological replicates was performed by Simon Obermeyer.



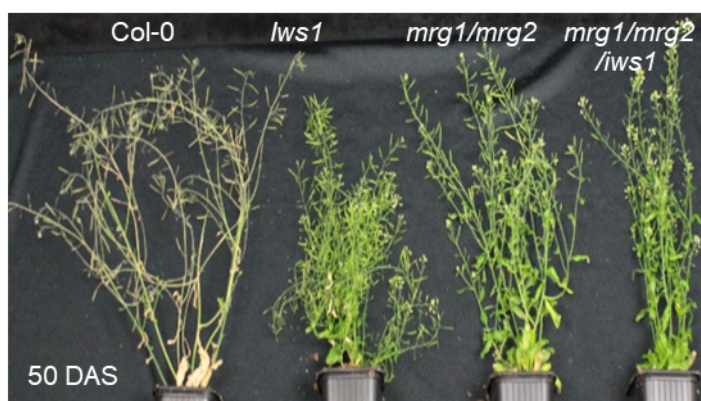




**Supplementary Figure 6. Size distribution of adapter-ligated fragment library.** The picks around 300bp represent the distribution of our samples after adaptor ligation and PCR amplification. The number on the peaks indicate the concentration in ng/ul. In green upper mark and in violet the lower mark. The number of the sample on every figure, represent the index primer number.

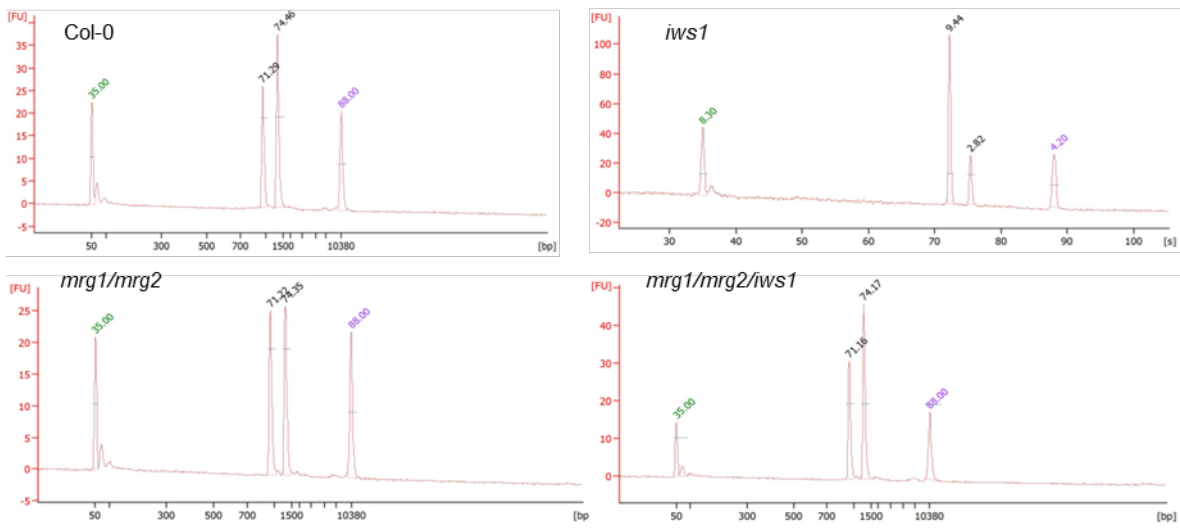


**Supplementary Figure 7** ChIP-seq tracks comparing RNAPII S2P and S5P reads over At1g68490 in Col-0, *spt4R3* and *iws1*. Y-axis represents the means of normalized reads. Heat maps were generated in Integrative Genomic Viewer (Robinson et al., 2011) using representative biological replicates. Gene model is depicted below mapped reads as follows: thin blue bars = UTRs; thick blue bars= exons; blue line= introns. Arrows indicate the AS event 5'SS. Reads are shown as vertical lines.

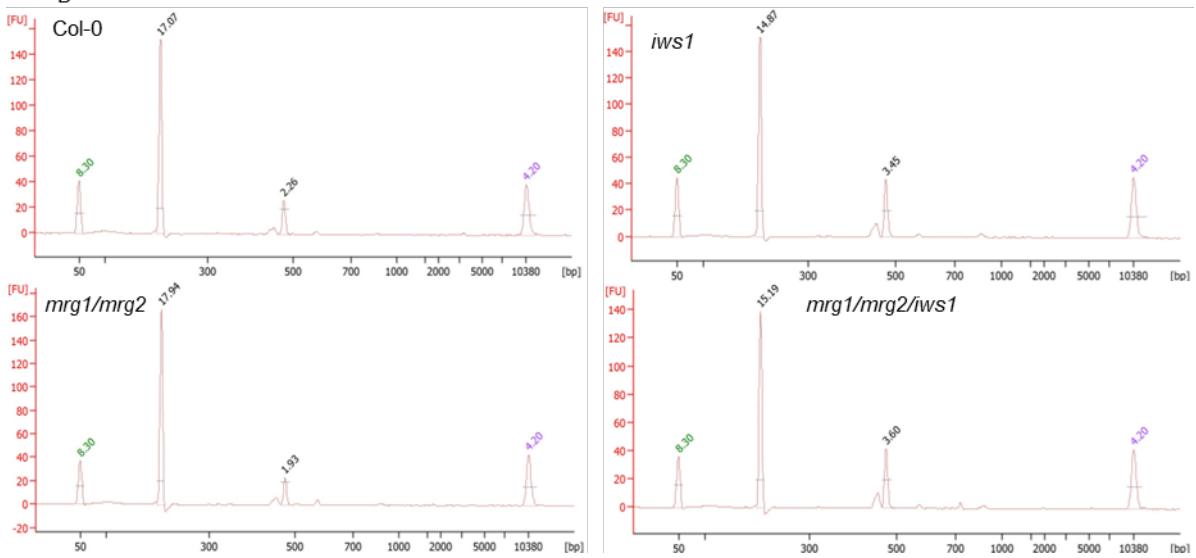


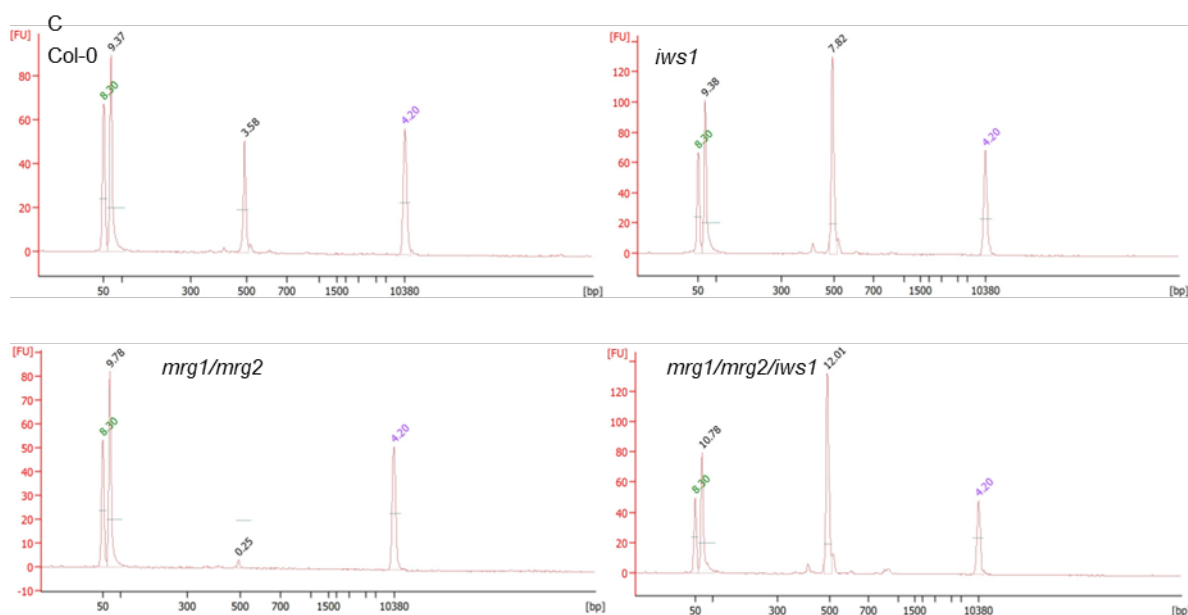
**Supplementary Figure 8.** Phenotypic analysis of Col-0, *iws1*, *mrg1/mrg2*, *mrg1/mrg2/iws1*. Phenotype of representative individuals is shown at Primary inflorescences 50 DAS.

A

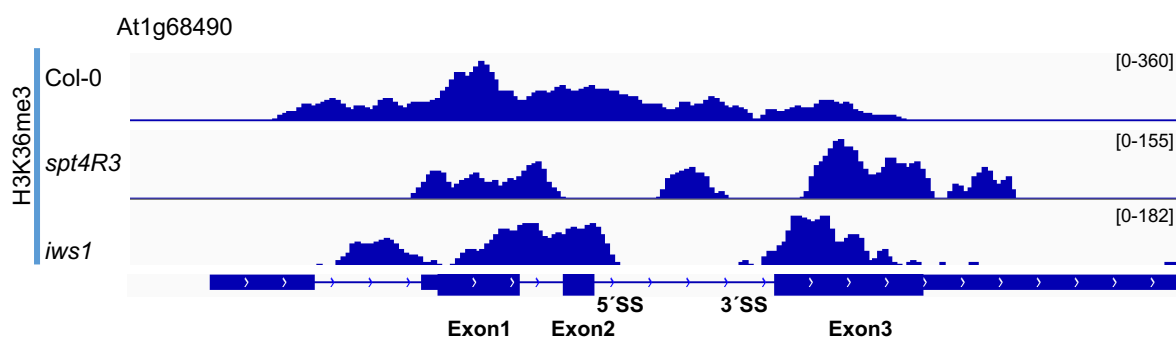


B





**Supplementary Figure 9. Quantitative analysis of the ratio of different splice variants in At1g09140 (SR30) (A), At3g58510 (B), At1g68490 (C).** Using RNA isolated of the different genotypes (Col-0, *iws1*, *mrg1/mrg2*, *mrg1/mrg2/iws1*). **At3g58510 (D)** (Col-0, *sdg7.2*, *sdg7.3*, *sdg8*). DNA fragments were amplified by RT-PCR with gene-specific primers spanning the transcripts. The obtained DNA fragments were analysed using a Bioanalyser. Numbers on top of the peaks give the retention time. The ratio between the Alternative Splicing (AS) and Constitutive splicing (CS) isoform was calculated dividing the nmol/l concentration of AS per nmol/l concentration of CS. Ratios were normalized to the corresponding control Col-0. The respective nmol/l concentration, used to calculate the splicing ratio, for every sample of every replicate are in the Supplementary table 14.



**Supplementary Figure 10. ChIP-seq tracks comparing H3K36me3 reads over At1g68490 in Col-0, *spt4R3* and *iws1*.** Y-axis represents the means of normalized reads. Heat maps were generated in Integrative Genomic Viewer (Robinson et al., 2011) using representative biological replicates. Gene model is depicted below mapped reads as follows: thin blue bars = UTRs; thick blue bars= exons; blue line= introns. Arrows indicate the AS event 5'SS. Reads are shown as vertical lines

## List of figures

<b>Figure 1.</b> RNAPII transcription cycle.....	3
<b>Figure 2.</b> Transcription elongation factors. ....	5
<b>Figure 3.</b> Splicing is regulated by cis- and trans-regulatory elements. ....	8
<b>Figure 4.</b> Schematic illustration of AS events. ....	10
<b>Figure 5.</b> AS can be explained by the kinetic model and the recruitment model. ....	13
<b>Figure 6.</b> MRG15-like proteins.....	16
<b>Figure 7.</b> Number of raw reads generated by Illumina sequencing in the analysed plant lines Col-0, <i>iws1</i> , <i>spt4R3</i> , <i>tflIs</i> and <i>tflIs/iws1</i> .....	20
<b>Figure 8.</b> AS events identified by RNAseq.....	21
<b>Figure 9.</b> Alternative splicing events differentially affected in <i>tef</i> mutants compared to Col-0....	22
<b>Figure 10.</b> Comparison of differentially alternatively spliced genes in <i>tef</i> mutants. ....	23
<b>Figure 11.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in the different <i>tef</i> single mutants. ....	24
<b>Figure 12.</b> RNAseq cover reads of the differentially alternatively spliced gene At3g58510. ....	25
<b>Figure 13.</b> The validation of 10 candidate genes differentially alternative spliced in at least one <i>tef</i> mutant according to RNAseq.....	27
<b>Figure 14.</b> Experimental validation of candidate genes identified by RNAseq to be differentially alternatively spliced. ....	28
<b>Figure 15.</b> Bioanalyzer analysis of a chromatin sample before and after sonication.....	30
<b>Figure 16.</b> Size distribution of adapter-ligated fragments in representative samples determined by Bioanalyzer measurement. ....	30
<b>Figure 17.</b> RNAPII Ser2P occupancy profile over all TAIR10 annotated genes determined by ChIP-seq.....	32
<b>Figure 18.</b> RNAPII Ser5P occupancy profile over all TAIR10 annotated genes determined by ChIP-seq.....	33
<b>Figure 19.</b> Mean RNAPII Ser2P distribution over exons and introns measured by ChIP-seq. ....	35
<b>Figure 20.</b> Mean RNAPII Ser5P distribution over exons and introns measured by ChIP-seq. ....	36
<b>Figure 21.</b> ChIP-qPCR validation experiment comparing the RNAPII distribution over At1g68490 in Col-0 and <i>spt4R3</i> . ....	38
<b>Figure 22.</b> ChIP-qPCR validation experiment comparing the RNAPII distribution over At1g68490 in Col-0 and <i>iws1</i> . ....	39
<b>Figure 23.</b> Motifs enriched around alternatively 5' and 3' splice sites.....	40
<b>Figure 24.</b> MRG1/2 and PTB2 affinity purification coupled with mass spectrometry. ....	42
<b>Figure 25.</b> MRG proteins and SDG7 are localized to the nucleus. ....	48

<b>Figure 26.</b> MRG1 and MRG2 protein directly interact with SDG7. ....	49
<b>Figure 27.</b> Protein interactions analysed by FRET experiments.....	50
<b>Figure 28.</b> Molecular characterization of <i>iws1-2</i> , <i>mrg1</i> and <i>mrg2</i> T-DNA lines. ....	52
<b>Figure 29.</b> Phenotypic analysis of <i>iws1 mrg1/mrg2</i> and <i>mrg1/mrg2/iws1</i> compared to Col-0. ...	53
<b>Figure 30.</b> Phenotypic analysis of <i>iws1 mrg1/mrg2</i> and <i>mrg1/mrg2/iws1</i> compared to Col-0. ...	54
<b>Figure 31.</b> AS of three selected genes in <i>iws1</i> and <i>mrg</i> mutant plants.....	55
<b>Figure 32.</b> Western blot analysis of global H3K36me3 levels.....	56
<b>Figure 33.</b> H3K36me3 density over all TAIR 10 annotated genes determined by ChIPseq.....	58
<b>Figure 34.</b> Mean H3K36me3 distribution over exons and introns measured by ChIP-seq. ....	59
<b>Figure 35.</b> ChIP-qPCR validation experiment comparing the H3K36me3 distribution over At1g68490 in Col-0 and <i>spt4R3</i> . ....	60
<b>Figure 36.</b> ChIP-qPCR validation experiment comparing the H3K36me3 distribution over At1g68490 in Col-0 and <i>iws1</i> .....	61
<b>Figure 37.</b> MRG2 is localized in the nucleus. ....	62
<b>Figure 38.</b> qPCR of At3g58510 after Chip experiment. ....	64
<b>Figure 39.</b> Splicing regulation through the IWS1-SETD2-MRG15-PTB2 adapter complex in human cells. ....	75
<b>Figure 40.</b> Model how the epigenetic mark H3K36me3 may affect the alternative splicing decision in Arabidopsis. ....	78
<b>Supplementary figure 1.</b> RNAseq cover reads of differentially alternatively spliced genes in Col-0 and the mutant(s) that show altered AS events. ....	131
<b>Supplementary figure 2.</b> Alternative Splicing isoform sequences.....	132
<b>Supplementary Figure 3.</b> Quantitative analysis of the ratio of different splice variants in At3g58510 (A), At1g68490 (B) At3g04500 (C), At3g61860 (RS31) (D), At2g25970 (E), At1g02840 (SR34) (F), At3g22210 (G), At1g50902 (H), At4g38280 (I), At2g04039 (J).....	139
<b>Supplementary figure 4.</b> Number of reads generated by Illumina Sequencing in the analysed sample Col-0, <i>spt4R3</i> , <i>iws1</i> with H3K36me3, S2P and S5P antibody (Mio. reads per sample). ....	140
<b>Supplementary Figure 5.</b> Principal Component Analysis (PCA) and Spearman correlation of datasets obtained by ChIP-seq for individual biological replicates.....	141
<b>Supplementary Figure 6.</b> Size distribution of adapter-ligated fragment library. ....	143
<b>Supplementary Figure 7.</b> ChIP-seq tracks comparing RNAPII S2P and S5P reads over At1g68490 in Col-0, <i>spt4R3</i> and <i>iws1</i> . ....	144
<b>Supplementary Figure 8.</b> Phenotypic analysis of Col-0, <i>iws1</i> , <i>mrg1/mrg2</i> , <i>mrg1/mrg2/iws1</i> . .	144
<b>Supplementary Figure 9.</b> Quantitative analysis of the ratio of different splice variants in At1g09140 (SR30) (A), At3g58510 (B), At1g68490 (C). ....	146
<b>Supplementary Figure 10.</b> ChIP-seq tracks comparing H3K36me3 reads over At1g68490 in Col-0, <i>spt4R3</i> and <i>iws1</i> . ....	146

## List of tables

<b>Table 1.</b> Candidate genes used for experimental validation. ....	29
<b>Table 2.</b> Splicing-related proteins co-purifying with PTB2-SG. ....	43
<b>Table 3.</b> Splicing-related proteins co-purifying with GS-MRG1. ....	45
<b>Table 4.</b> Splicing-related proteins co-purifying with GS-MRG2. ....	46
<b>Table 5.</b> Selected Proteins co-purifying with GS-MRG2 (AP without benzonase). ....	47
<b>Table 6.</b> SR proteins identified in the different <i>tefs</i> to be differentially alternatively spliced. ....	67
<b>Table 7.</b> Kits and instruments used in this study. ....	78
<b>Table 8.</b> List of oligonucleotides for RT-PCR, qPCR and genotyping. ....	80
<b>Table 9.</b> List of oligonucleotides for cloning. ....	83
<b>Table 10.</b> Bacterial and Yeast strain. ....	84
<b>Table 11.</b> List of vectors created in the course of this study. from this study. ....	84
<b>Table 12.</b> List of T-DNA and RNAi lines. ....	85
<b>Table 13.</b> Databases, Online Tools and Software. ....	85
<b>Table 14.</b> PCR cycle programs for Taq, KAPA HiFi and Hercules II. ....	88
<b>Table 15.</b> Reagents used for Taq, KAPA HiFi and Hercules II. ....	88
<b>Table 16.</b> qRT-PCR cycle program. ....	89
<b>Table 17.</b> Buffers used for ChIP. ....	93
<b>Supplementary table 1.</b> Differentially alternatively spliced genes in <i>tef</i> mutants. ....	103
<b>Supplementary table 2.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>spt4R3</i> involved in metabolic process. ....	107
<b>Supplementary table 3.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>spt4R3</i> involved in cell division, development and biogenesis. ....	108
<b>Supplementary table 4.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>spt4R3</i> involved in RNA splicing, RNA processing, Stress response and Light response. ....	108
<b>Supplementary table 5.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>tflls</i> involved in metabolic process. ....	109
<b>Supplementary table 6.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>tflls</i> involved in cell division, development and biogenesis. ....	110
<b>Supplementary table 7.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>tflls</i> involved in RNA processing, Stress response and Light response. ....	110
<b>Supplementary table 8.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>iws1</i> involved in metabolic process. ....	111

<b>Supplementary table 9.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>iws1</i> involved in cell division, development and biogenesis .....	114
<b>Supplementary table 10.</b> Gene ontology enrichment analysis of genes differentially alternatively spliced in <i>iws1</i> involved in RNA splicing, RNA processing, Stress response and Light response. ....	115
<b>Supplementary Table 11.</b> Alternative splice forms were co-amplified and quantified using a Bioanalyzer.....	115
<b>Supplementary table 12.</b> ChIP seq sample information. ....	117
<b>Supplementary table 13.</b> Phenotypic analysis of <i>mrg1/mrg2/iws1</i> compared to the double mutant <i>mrg1/mrg2</i> , single mutant <i>iws1</i> and Col-0. ....	117
<b>Supplementary table 14.</b> Phenotypic analysis of <i>mrg1/mrg2/iws1</i> compared to the double mutant <i>mrg1/mrg2</i> , single mutant <i>iws1</i> and Col-0. ....	117
<b>Supplementary table 15.</b> Transcription-related proteins co-purifying with PTB2-SG.....	118
<b>Supplementary table 16.</b> Transcription-related proteins co-purifying with GS-MRG1.....	118
<b>Supplementary table 17.</b> Transcription-related proteins co-purifying with GS-MRG2.....	118
<b>Supplementary table 18.</b> List of interactors co-purifying with GS-MRG2, MRG2-SG, GS-MRG1, MRG1-SG, MRG2 without benzonase, PTB2-GS. ....	119
<b>Supplementary Table 19.</b> Alternative splice forms were coamplified and quantified using a Bioanalyzer.....	129



## Publications

**Sörensen BB., Ehrnsberger HF., Esposito S., Pfab A., Bruckmann A., Hauptmann J., Meister G., Merkl R., Schubert T., Längst G., Melzer M., Grasser M., Grasser KD.** (2017). The Arabidopsis THO/TREX component TEX1 functionally interacts with MOS11 and modulates mRNA export and alternative splicing events. *Plant Mol Biol.*, 93(3):283-298

**Ruggiero M., Fiore MG., Magherini S., Esposito S., Moroni G., Gulisano M., Pacini S.** (2012) Transcranial sonography in the diagnosis, follow-up and treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of lime*, 6(1):23-28



## Bibliography

**Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T.** (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell.*

**Albaqami, M., and Reddy, A.S.N.** (2018). Development of an in vitro pre-mRNA splicing assay using plant nuclear extract. *Plant Methods.*

**Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D.** (2010). Splicing-Dependent RNA polymerase pausing in yeast. *Mol. Cell.*

**Allemand, E., Batsché, E., and Muchardt, C.** (2008). Splicing, transcription, and chromatin: a ménage à trois. *Curr. Opin. Genet. Dev.*

**Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J.** (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.*

**Antosz, W., Pfab, A., Ehrnsberger, H.F., Holzinger, P., Köllen, K., Mortensen, S.A., Bruckmann, A., Schubert, T., Längst, G., Griesenbeck, J., et al.** (2017). The Composition of the Arabidopsis RNA Polymerase II Transcript Elongation Complex Reveals the Interplay between Elongation and mRNA Processing Factors. *Plant Cell.* 29, 854–870.

**Antosz, W., Deforges, J., Begcy, K., Bruckmann, A., Poirier, Y., Dresselhaus, T., and Grasser, K.D.** (2020). Critical Role of Transcript Cleavage in Arabidopsis RNA Polymerase II Transcriptional Elongation. *Plant Cell.*

**Barbazuk, W.B., Fu, Y., and McGinnis, K.M.** (2008). Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Res.*

**Barta, A., Kalyna, M., and Lorković, Z.J.** (2008). Plant SR proteins and their functions. *Curr. Top. Microbiol. Immunol.*

**Bentley, D.L.** (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175.

**Bentsink, L., Jowett, J., Hanhart, C.J., and Koornneef, M.** (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.*

**Berget, S.M., Moore, C., and Sharp, P.A.** (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci.*

**Bertram, M.J., and Pereira-Smith, O.M.** (2001). Conservation of the MORF4 related gene family: Identification of a new chromo domain subfamily and novel protein motif. *Gene.*

- Beyer, A.L., and Osheim, Y.N.** (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*
- Bourgeois, C.F., Lejeune, F., and Stévenin, J.** (2004). Broad Specificity of SR (Serine{plus 45 degree rule}Arginine) Proteins in the Regulation of Alternative Splicing of Pre-Messenger RNA. *Prog. Nucleic Acid Res. Mol. Biol.*
- Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y.** (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.*
- Brown, J.W.S., and Simpson, C.G.** (1998). **SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.*
- Bu, Z., Yu, Y., Li, Z., Liu, Y., Jiang, W., Huang, Y., and Dong, A.W.** (2014). Regulation of Arabidopsis Flowering by the Histone Mark Readers MRG1/2 via Interaction with CONSTANS to Modulate FT Expression. *PLoS Genet.* 10, 1–11.
- Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M.** (2005). Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.*
- Cao, Y., Wen, L., Wang, Z., and Ma, L.** (2015). SKIP Interacts with the Paf1 Complex to Regulate Flowering via the Activation of FLC Transcription in Arabidopsis. *Mol. Plant.*
- Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M.** (2010). Global analysis of nascent rna reveals transcriptional pausing in terminal exons. *Mol. Cell.*
- Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M.** (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* 165, 372–381.
- Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., et al.** (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell.*
- Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A.S.N., Byrne, L.J., Wilson, C.M., and Syed, N.H.** (2019). Alternative splicing and protein diversity: Plants versus animals. *Front. Plant Sci.*
- Chen, W., Luo, L., and Zhang, L.** (2010). The organization of nucleosomes around splice sites. *Nucleic Acids Res.*
- Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J.** (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell.*
- Chu, Y., Simic, R., Warner, M.H., Arndt, K.M., and Prelich, G.** (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J.*
- Clancy, S.** (2008). RNA transcription by RNA polymerase: pro-karyotes vs eukaryotes. *Lett. To Nat.*

- De Conti, L., Baralle, M., and Buratti, E.** (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA*.
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R.** (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol. Cell*.
- Das, U., Nguyen, H., and Xie, J. (2019). Transcriptome protection by the expanded family of hnRNPs. *RNA Biol*.
- Ding, Y., Avramova, Z., and Fromma, M.** (2011). Two distinct roles of Arabidopsis homolog of trithorax1 (ATX1) at promoters and within transcribed regions of ATX1-regulated genes. *Plant Cell*.
- Dolata, J., Guo, Y., Kołowerzo, A., Smoliński, D., Brzyżek, G., Jarmołowski, A., and Świeżewski, S.** (2015). NTR 1 is required for transcription elongation checkpoints at alternative exons in Arabidopsis. *EMBO J.* 34, 544–558.
- Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Ratsch, G., and Wachter, A.** (2013). Nonsense-Mediated Decay of Alternative Precursor mRNA Splicing Variants Is a Major Determinant of the Arabidopsis Steady State Transcriptome. *Plant Cell* 25, 3726–3742.
- Drexler, H.L., Choquet, K., and Churchman, L.S.** (2019). Human co-transcriptional splicing kinetics and coordination revealed by direct nascent RNA sequencing. *BioRxiv*.
- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Muñoz, M.J., Le Jossic-Corcós, C., Corcos, L., and Kornblihtt, A.R.** (2014). How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. *Mol. Cell*.
- Dürr, J., Lolas, I.B., Sørensen, B.B., Schubert, V., Houben, A., Melzer, M., Deutzmann, R., Grasser, M., and Grasser, K.D.** (2014). The transcript elongation factor SPT4/SPT5 is involved in auxin-related gene expression in Arabidopsis. *Nucleic Acids Res.* 42, 4332–4347.
- Dvinge, H.** (2018). Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett*.
- Edwards, K., Johnstone, C., and Thompson, C.** (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.*
- Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends Genet.*
- Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.I.** (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* (80- .).
- Ehrnsberger, H.F., Pfaff, C., Hachani, I., Flores-Tornero, M., Soerensen, B.B., Längst, G., Sprunck, S., Grasser, M., and Grasser, K.D.** (2019). The UAP56-interacting export factors UIEF1 and UIEF2 function in mRNA export. *Plant Physiol*.

- Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J., and Schaal, H.** (2013). Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA*.
- Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., Yan, S., Zhang, W., He, H., Ma, L., et al.** (2015). SKIP Confers Osmotic Tolerance during Salt Stress by Controlling Alternative Gene Splicing in Arabidopsis. *Mol. Plant*.
- Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C.** (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res*.
- Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G., Reddy, A.S.N., Megraw, M., and Mockler, T.C.** (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant*.
- Fish, R.N., and Kane, C.M.** (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta - Gene Struct. Expr.*
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L.** (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev*.
- Fusby, B., Kim, S., Erickson, B., Kim, H., Peterson, M.L., and Bentley, D.L.** (2015). Coordination of RNA Polymerase II Pausing and 3' end processing factor recruitment with alternative polyadenylation. *Mol. Cell. Biol.*
- Galej, W.P.** (2018). Structural studies of the spliceosome: Past, present and future perspectives. *Biochem. Soc. Trans.*
- Gallego-Paez, L.M., Bordone, M.C., Leote, A.C., Saraiva-Agostinho, N., Ascensão-Ferreira, M., and Barbosa-Morais, N.L.** (2017). Alternative splicing: the pledge, the turn, and the prestige: The key role of alternative splicing in human biological systems. *Hum. Genet.*
- Gelfman, S., Cohen, N., Yearim, A., and Ast, G.** (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res*.
- Godoy Herz, M.A., Kubaczka, M.G., Brzyżek, G., Servi, L., Krzyszton, M., Simpson, C., Brown, J., Swiezewski, S., Petrillo, E., and Kornblihtt, A.R.** (2019). Light Regulates Plant Alternative Splicing through the Control of Transcriptional Elongation. *Mol. Cell* 1–9.
- Grasser, M., and Grasser, K.D.** (2018). The plant RNA polymerase II elongation complex: A hub coordinating transcript elongation and mRNA processing. *Transcription*.
- Grasser, M., Kane, C.M., Merkle, T., Melzer, M., Emmersen, J., and Grasser, K.D.** (2009). Transcript Elongation Factor TFIIS Is Involved in Arabidopsis Seed Dormancy. *J. Mol. Biol.*
- Graveley, B.R.** (2001). Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.*

- Gu, B., Eick, D., and Bensaude, O.** (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.*
- Hajheidari, M., Koncz, C., and Eick, D. (2013). Emerging roles for RNA polymerase II CTD in Arabidopsis. *Trends Plant Sci.*
- Hall, J.A., and Georgel, P.T.** (2011). **RNA Processing (InTech).**
- Hammer, Ø., Harper, D., and Ryan, P.** (2001). Past: Paleontological statistics software package for education and data analysis. *Paleontol. Electron.*
- Harlen, K.M., and Churchman, L.S.** (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.*
- Hartmann, L., Drewe-Boß, P., Wießner, T., Wagner, G., Geue, S., Lee, H.-C., Obermüller, D.M., Kahles, A., Behr, J., Sinz, F.H., et al.** (2016a). LARGE-SCALE BIOLOGY ARTICLE Alternative Splicing Substantially Diversifies the Transcriptome during Early Photomorphogenesis and Correlates with the Energy Availability in Arabidopsis.
- Hartmann, L., Drewe-Boß, P., Wießner, T., Wagner, G., Geue, S., Lee, H.-C., Obermüller, D.M., Kahles, A., Behr, J., Sinz, F.H., et al.** (2016b). Alternative Splicing Substantially Diversifies the Transcriptome during Early Photomorphogenesis and Correlates with the Energy Availability in Arabidopsis. *Plant Cell* 28, 2715–2734.
- Hartzog, G.A., and Fu, J.** (2013). The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.*
- Haseloff, J., Siemering, K.R., Prasher, D.C., and Hodge, S.** (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc. Natl. Acad. Sci. U. S. A.*
- He, Y., Doyle, M.R., and Amasino, R.M.** (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes Dev.*
- Heidemann, M., and Eick, D.** (2012). Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code. *RNA Biol.*
- Heidemann, M., Hintermair, C., Voß, K., and Eick, D.** (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.*
- Hetzl, J., Duttke, S.H., Benner, C., and Chory, J.** (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12316–12321.
- Hirtreiter, A., Damsma, G.E., Cheung, A.C.M., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C.R., Cramer, P., and Werner, F.** (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.*
- Hogg, R., McGrail, J.C., and O’Keefe, R.T.** (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.*

- Howe, K.J., Kane, C.M., and Ares, M.** (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA*.
- Huang, H., Yu, S., Liu, H., and Sun, X.** (2012). Nucleosome organization in sequences of alternative events in human genome. *BioSystems*.
- Iwabuchi, K., Li, B., Bartel, P., and Fields, S.** (1993). Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. *Oncogene*.
- Izquierdo, J.M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., Bilbao, D., and Valcárcel, J.** (2005). Regulation of fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell*.
- Jabre, I., Reddy, A.S.N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L.J., Wilson, C.M., and Syed, N.H.** (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.*
- Jangi, M., and Sharp, P.A.** (2014). Building robust transcriptomes with master splicing factors. *Cell*.
- Jeong, S.** (2017). SR proteins: Binders, regulators, and connectors of RNA. *Mol. Cells*.
- Jeronimo, C., Collin, P., and Robert, F. (2016). The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *J. Mol. Biol.*
- Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D.** (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell*.
- Jia, T., Zhang, B., You, C., Zhang, Y., Zeng, L., Li, S., Johnson, K.C.M., Yu, B., Li, X., and Chen, X.** (2017). The arabidopsis MOS4-associated complex promotes microRNA biogenesis and precursor messenger RNA splicing. *Plant Cell*.
- Jimeno-González, S., Payán-Bravo, L., Muñoz-Cabello, A.M., Guijo, M., Gutierrez, G., Prado, F., and Reyes, J.C.** (2015). Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.*
- Jonkers, I., and Lis, J.T.** (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.*
- Jonkers, I., Kwak, H., and Lis, J.T.** (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 3.
- Joshi, A.A., and Struhl, K.** (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation. *Mol. Cell*.
- Jurica, M.S., and Moore, M.J.** (2003). Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell*.
- Kanno, T., Venhuizen, P., Wu, M.-T., Chiou, P., Chang, C.-L., Kalyna, M., Matzke, A.J.M., and Matzke, M.** (2020). A Collection of Pre-mRNA Splicing Mutants in *Arabidopsis thaliana*. *G3&#58; Genes|Genomes|Genetics* g3.400998.2019.



- Kettenberger, H., Armache, K.J., and Cramer, P.** (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J.** (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381.
- Komarnitsky, P., Cho, E.J., and Buratowski, S.** (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*
- Kornblihtt, A.R., De La Mata, M., Fededa, J.P., Muñoz, M.J., and Nogués, G.** (2004). Multiple links between transcription and splicing. *RNA*.
- Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F.** (2002). RNA Polymerase II Elongation Factors of *Saccharomyces cerevisiae*: a Targeted Proteomics Approach. *Mol. Cell. Biol.*
- Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. *Annu. Rev. Genet.*
- De La Mata, M., and Kornblihtt, A.R.** (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.*
- De La Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R.** (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell*.
- Labadorf, A., Link, A., Rogers, M.F., Thomas, J., Reddy, A.S.N., and Ben-Hur, A.** (2010). Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. *BMC Genomics*.
- Lareau, L.F., Brooks, A.N., Soergel, D.A.W., Meng, Q., and Brenner, S.E.** (2007). The Coupling of Alternative Splicing and Nonsense-Mediated mRNA Decay. In *Proceedings of the National Academy of Sciences*, pp. 190–211.
- Lee, G., and Blenis, J.** (2014). Akt-ivation of RNA Splicing. *Mol. Cell* 53, 519–520.
- Lee, J.S., and Shilatifard, A.** (2007). A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*
- Lee, J., Yun, J.-Y., Zhao, W., Shen, W.-H., and Amasino, R.M.** (2015). A methyltransferase required for proper timing of the vernalization response in *Arabidopsis*. *Proc. Natl. Acad. Sci.*
- Lee, Y., Park, D., and Iyer, V.R.** (2017). The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. *Nucleic Acids Res.*
- Lenasi, T., and Barboric, M.** (2013). Mutual relationships between transcription and pre-mRNA processing in the synthesis of mRNA. *Wiley Interdiscip. Rev. RNA*.
- Li, L., Ye, H., Guo, H., and Yin, Y.** (2010). *Arabidopsis* IWS1 interacts with transcription factor BES1 and is involved in plant steroid hormone brassinosteroid regulated gene expression. *Proc. Natl. Acad. Sci.*

- Li, S., Wang, Y., Zhao, Y., Zhao, X., Chen, X., and Gong, Z.** (2020). Global Co-transcriptional Splicing in Arabidopsis and the Correlation with Splicing Regulation in Mature RNAs. *Mol. Plant*.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., and Hartzog, G.A.** (2003). Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Mol. Cell. Biol.*
- Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S.** (2004). Two Cyclin-Dependent Kinases Promote RNA Polymerase II Transcription and Formation of the Scaffold Complex. *Mol. Cell. Biol.*
- Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S.** (2009). Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. *Mol. Cell. Biol.*
- Liu, Y., Geyer, R., van Zanten, M., Carles, A., Li, Y., Hörold, A., van Nocker, S., and Soppe, W.J.J.** (2011). Identification of the Arabidopsis reduced dormancy 2 gene uncovers a role for the polymerase associated factor 1 complex in seed dormancy. *PLoS One*.
- Liu, Y., Wu, H., Yu, Y., and Huang, Y.** (2016). Structural studies on MRG701 chromodomain reveal a novel dimerization interface of MRG proteins in green plants. *Protein Cell*.
- Liu, Y., González-Porta, M., Santos, S., Brazma, A., Marioni, J.C., Aebersold, R., Venkitaraman, A.R., and Wickramasinghe, V.O.** (2017). Impact of Alternative Splicing on the Human Proteome. *Cell Rep*.
- Loomis, R.J., Naoe, Y., Parker, J.B., Savic, V., Bozovsky, M.R., Macfarlan, T., Manley, J.L., and Chakravarti, D.** (2009). Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation. *Mol. Cell*.
- Lorković, Z.J., Lehner, R., Forstner, C., and Barta, A.** (2005). Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA*.
- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T.** (2010). Regulation of alternative splicing by histone modifications. *Science* (80-. ).
- Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T.** (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*.
- Mancini, E., Sanchez, S.E., Romanowski, A., Schlaen, R.G., Sanchez-Lamas, M., Cerdán, P.D., and Yanovsky, M.J.** (2016). Acute Effects of Light on Alternative Splicing in Light-Grown Plants. *Photochem. Photobiol.* 92, 126–133.
- Marquez, Y., Brown, J.W.S., Simpson, C., Barta, A., and Kalyna, M.** (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res*.
- Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M.** (2015). Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res*.

- Marshall, C.M., Tartaglio, V., Duarte, M., and Harmon, F.G.** (2016). The Arabidopsis sickle Mutant Exhibits Altered Circadian Clock Responses to Cool Temperatures and Temperature-Dependent Alternative Splicing . *Plant Cell*.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J.-F., Revil, T., and Chabot, B.** (2007). hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.*
- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P.** (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 17, 1272–1278.
- Mayer, A., Di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S.** (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*.
- Mayer, A., Landry, H.M., and Churchman, L.S.** (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.*
- Mortensen, S.A., and Grasser, K.D.** (2014). The seed dormancy defect of Arabidopsis mutants lacking the transcript elongation factor TFIIIS is caused by reduced expression of the DOG1 gene. *FEBS Lett.*
- Morton, M., AlTamimi, N., Butt, H., Reddy, A.S.N., and Mahfouz, M.** (2019). Serine/Arginine-rich protein family of splicing regulators: New approaches to study splice isoform functions. *Plant Sci.* 283, 127–134.
- Muhammad, T., Zhang, F., Zhang, Y., and Liang, Y.** (2019). RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. *Cells* 8, 38.
- Naftelberg, S., Schor, I.E., Ast, G., and Kornblihtt, A.R.** (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Annu. Rev. Biochem.* 84, 165–198.
- Nechaev, S., Fargo, D.C., Santos, G. Dos, Liu, L., Gao, Y., and Adelman, K.** (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* (80-. ).
- Nishida, S., Kakei, Y., Shimada, Y., and Fujiwara, T.** (2017). Genome-wide analysis of specific alterations in transcript structure and accumulation caused by nutrient deficiencies in *Arabidopsis thaliana*. *Plant J.*
- Nogués, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A.R.** (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.*
- Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J.** (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*.

- Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M.** (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell*.
- Oqani, R.K., Lin, T., Lee, J.E., Kang, J.W., Shin, H.Y., and Il Jin, D.** (2019). lws1 and Spt6 Regulate Trimethylation of Histone H3 on Lysine 36 through Akt Signaling and are Essential for Mouse Embryonic Genome Activation. *Sci. Rep.*
- Pajoro, A., Severing, E., Angenent, G.C., and Immink, R.G.H.** (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.*
- Palusa, S.G., and Reddy, A.S.N.** (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytol.*
- Palusa, S.G., Ali, G.S., and Reddy, A.S.N.** (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: Regulation by hormones and stresses. *Plant J.*
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J.** (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*
- Peng, M., Li, Z., Zhou, N., Ma, M., Jiang, Y., Dong, A., Shen, W.-H., and Li, L.** (2018). Linking PHYTOCHROME-INTERACTING FACTOR to Histone Modification in Plant Shade Avoidance. *Plant Physiol.*
- Peterlin, B.M., and Price, D.H.** (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Mol. Cell*.
- Petrillo, E., Godoy Herz, M.A., Fuchs, A., Reifer, D., Fuller, J., Yanovsky, M.J., Simpson, C., Brown, J.W.S., Barta, A., Kalyna, M., et al.** (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* (80- ).
- Pfab, A., Antosz, W., Holzinger, P., Bruckmann, A., Griesenbeck, J., and Grasser, K.D.** (2017). Analysis of in vivo chromatin and protein interactions of Arabidopsis transcript elongation factors. In *Methods in Molecular Biology*, p.
- Pfaff, C., Ehrnsberger, H.F., Flores-Tornero, M., Soerensen, B.B., Schubert, T., Längst, G., Griesenbeck, J., Sprunck, S., Grasser, M., and Grasser, K.D.** (2018). ALY RNA-binding proteins are required for nucleo-cytosolic mRNA transport and modulate plant growth and development. *Plant Physiol.*
- Pujari, V., Radebaugh, C.A., Chodaparambil, J. V., Muthurajan, U.M., Almeida, A.R., Fischbeck, J.A., Luger, K., and Stargell, L.A.** (2010). The transcription factor Spn1 regulates gene expression via a highly conserved novel structural motif. *J. Mol. Biol.*
- Rayson, S., Arciga-Reyes, L., Wootton, L., de Torres Zabala, M., Truman, W., Graham, N., Grant, M., and Davies, B.** (2012). A role for nonsense-mediated mRNA decay in plants: Pathogen responses are induced in Arabidopsis thaliana nmd mutants. *PLoS One*.

- Reddy, A.S.N.** (2007). Alternative Splicing of Pre-Messenger RNAs in Plants in the Genomic Era. *Annu. Rev. Plant Biol.* 58, 267–294.
- Reddy, A.S.N., and Shad Ali, G.** (2011). Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses. *Wiley Interdiscip. Rev. RNA* 2, 875–889.
- Reddy, A.S.N., Marquez, Y., Kalyna, M., and Barta, A.** (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell*.
- Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., et al.** (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.*
- Ruhl, C., Stauffer, E., Kahles, A., Wagner, G., Drechsel, G., Ratsch, G., and Wachter, A.** (2012). Polypyrimidine Tract Binding Protein Homologs from Arabidopsis Are Key Regulators of Alternative Splicing with Implications in Fundamental Developmental Processes. *Plant Cell* 24, 4360–4375.
- Sahr, T., Adam, T., Fizames, C., Maurel, C., and Santoni, V.** (2010). O-carboxyl- and N-methyltransferases active on plant aquaporins. *Plant Cell Physiol.* 51, 2092–2104.
- Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L.** (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J. Mol. Biol.*
- Sanidas, I., Polytarchou, C., Hatziapostolou, M., Ezell, S.A., Kottakis, F., Hu, L., Guo, A., Xie, J., Comb, M.J., Iliopoulos, D., et al.** (2014). Phosphoproteomics Screen Reveals Akt Isoform-Specific Signals Linking RNA Processing to Lung Cancer. *Mol. Cell* 53, 577–590.
- Sapra, A.K., Änkö, M.L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.M., and Neugebauer, K.M.** (2009). SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. *Mol. Cell*.
- Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J.** (2006). The Polypyrimidine Tract Binding Protein (PTB) Represses Splicing of Exon 6B from the  $\beta$ -Tropomyosin Pre-mRNA by Directly Interfering with the Binding of the U2AF65 Subunit. *Mol. Cell. Biol.*
- Saunders, A., Core, L.J., and Lis, J.T.** (2006). Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.*
- Schwartz, S., Meshorer, E., and Ast, G.** (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.*
- Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q.** (2010). Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.*
- Seo, P.J., Ryu, J., Kang, S.K., and Park, C.M.** (2011). Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in Arabidopsis. *Plant J.*

- Sharma, S., Falick, A.M., and Black, D.L.** (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal e complex. *Mol. Cell*.
- Shaul, O.** (2015). Unique Aspects of Plant Nonsense-Mediated mRNA Decay. *Trends Plant Sci.* 20, 767–779.
- Shen, H., and Green, M.R.** (2004). A pathway of sequential Arginine-Serine-Rich Domain-Splicing signal interactions during mammalian spliceosome assembly. *Mol. Cell*.
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S.** (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*.
- Simpson, C.G., Jennings, S.N., Clark, G.P., Thow, G., and Brown, J.W.S.** (2004). Dual functionality of a plant U-rich intronic sequence element. *Plant J.*
- Simpson, C.G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., McNicol, J., Barta, A., and Brown, J.W.S.** (2008). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.*
- Sims, R.J., Belotserkovskaya, R., and Reinberg, D.** (2004). Elongation by RNA polymerase II: The short and long of it. *Genes Dev.*
- Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schüller, R., Drury, W.J., Eick, D., and Reinberg, D.** (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* (80- ).
- Sorenson, M.R., Jha, D.K., Ucles, S.A., Flood, D.M., Strahl, B.D., Stevens, S.W., and Kress, T.L.** (2016). Histone H3K36 methylation regulates pre-mRNA splicing in *Saccharomyces cerevisiae*. *RNA Biol.*
- Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B.** (2009). Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Mol. Cell* 36, 245–254.
- Staiger, D., Zecca, L., Wieczorek Kirk, D.A., Apel, K., and Eckstein, L.** (2003). The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.* 33, 361–371.
- Stauffer, E., Westermann, A., Wagner, G., and Wachter, A.** (2010). Polypyrimidine tract-binding protein homologues from *Arabidopsis* underlie regulatory circuits based on alternative splicing and downstream control. *Plant J.* 64, 243–255.
- Subtil-Rodríguez, A., and Reyes, J.C.** (2011). To cross or not to cross the nucleosome, that is the elongation question... *RNA Biol.*
- Svejstrup, J.Q.** (2007). Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem. Sci.* 32, 165–171.

- Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W.S.** (2012). Alternative splicing in plants - coming of age. *Trends Plant Sci.*
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R.** (2009). Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.*
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R.** (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*
- Újvári, A., and Luse, D.S.** (2004). Newly initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *J. Biol. Chem.*
- Ullah, F., Hamilton, M., Reddy, A.S.N., and Ben-Hur, A.** (2018). Exploring the relationship between intron retention and chromatin accessibility in plants. *BMC Genomics.*
- Van Leene, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Geerinck, J., Van Isterdael, G., Witters, E., and De Jaeger, G.** (2011). Isolation of transcription factor complexes from Arabidopsis cell suspension cultures by tandem affinity purification. *Methods Mol. Biol.*
- Van Leene, J., Eeckhout, D., Cannoot, B., De Winne, N., Persiau, G., Van De Slijke, E., Vercruyse, L., Dedecker, M., Verkest, A., Vandepoele, K., et al.** (2015). An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. *Nat. Protoc.*
- Van Lijsebettens, M., and Grasser, K.D.** (2014). Transcript elongation factors: Shaping transcriptomes after transcript initiation. *Trends Plant Sci.*
- Van Lijsebettens, M., Dürr, J., Woloszynska, M., and Grasser, K.D.** (2014). Elongator and SPT4/SPT5 complexes as proxy to study RNA polymerase II transcript elongation control of plant development. *Proteomics.*
- Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A.E., Schedl, P., and Tyagi, S.** (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell.*
- Verhage, L., Severing, E.I., Bucher, J., Lammers, M., Busscher-Lange, J., Bonnema, G., Rodenburg, N., Proveniers, M.C.G., Angenent, G.C., and Immink, R.G.H.** (2017). Splicing-related genes are alternatively spliced upon changes in ambient temperatures in plants. *PLoS One.*
- Vinayachandran, V., Reja, R., Rossi, M.J., Park, B., Rieber, L., Mittal, C., Mahony, S., and Pugh, B.F.** (2018). Widespread and precise reprogramming of yeast protein-genome interactions in response to heat shock. *Genome Res.*
- Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P.** (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature.*
- Wachter, A., Rühl, C., and Stauffer, E.** (2012). The role of polypyrimidine tract-binding proteins and other hnRNP proteins in plant splicing regulation. *Front. Plant Sci.*

- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al.** (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.*
- Wagner, E.J., and Carpenter, P.B.** (2012). Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.*
- Wallace, E.W.J., and Beggs, J.D.** (2017). Extremely fast and incredibly close: Cotranscriptional splicing in budding yeast. *RNA.*
- Wang, B.B., and Brendel, V.** (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7175–7180.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B.** (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature.*
- Wang, X., Wu, F., Xie, Q., Wang, H., Wang, Y., Yue, Y., Gahura, O., Ma, S., Liu, L., Cao, Y., et al.** (2012). SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in *Arabidopsis*. *Plant Cell.*
- Wang, X., Chen, J., Xie, Z., Liu, S., Nolan, T., Ye, H., Zhang, M., Guo, H., Schnable, P.S., Li, Z., et al.** (2014). Histone lysine methyltransferase SDG8 is involved in brassinosteroid- regulated gene expression in *Arabidopsis thaliana*. *Mol. Plant* 7, 1303–1315.
- Weber, C.M., Ramachandran, S., and Henikoff, S.** (2014). Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase. *Mol. Cell.*
- Widiez, T., El Kafafi, E.S., Girin, T., Berr, A., Ruffel, S., Krouk, G., Vayssières, A., Shen, W.H., Coruzzi, G.M., Gojon, A., et al.** (2011). High Nitrogen Insensitive 9 (HNI9)-mediated systemic repression of root NO<sub>3</sub>- uptake is associated with changes in histone methylation. *Proc. Natl. Acad. Sci. U. S. A.*
- Will, C.L., and Lührmann, R.** (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*
- Wu, Z., Zhu, D.D., Lin, X., Miao, J., Gu, L., Deng, X., Yang, Q., Sun, K., Zhu, D.D., Cao, X., et al.** (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Cell* 161, 1–15.
- Xie, T., Zmyslowski, A.M., Zhang, Y., and Radhakrishnan, I.** (2015). Structural Basis for Multi-specificity of MRG Domains. *Structure.*
- Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.-P., Steinmetz, A., and Shen, W.-H.** (2007). Di- and Tri- but Not Monomethylation on Histone H3 Lysine 36 Marks Active Transcription of Genes Involved in Flowering Time Regulation and Other Processes in *Arabidopsis thaliana*. *Mol. Cell. Biol.*



- Xu, Y., Gan, E.-S., Zhou, J., Wee, W.-Y., Zhang, X., and Ito, T.** (2014). Arabidopsis MRG domain proteins bridge two histone modifications to elevate expression of flowering genes. *Nucleic Acids Res.* 42, 10960–10974.
- Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A.** (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct *lws1*-dependent mRNA splicing and export. *Genes Dev.*
- Yoh, S.M., Lucas, J.S., and Jones, K.A.** (2008). The *lws1*:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.*
- Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., Jones, K.A., Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., et al.** (2010). The structure of an *lws1*/Spt6 complex reveals an interaction domain conserved in TFIIIS, Elongin A and Med26. *Genes Dev.* 29, 3979–3991.
- Zhang, L., Fletcher, A.G.L., Cheung, V., Winston, F., and Stargell, L.A.** (2008). Spn1 Regulates the Recruitment of Spt6 and the Swi/Snf Complex during Transcriptional Activation by RNA Polymerase II. *Mol. Cell. Biol.*
- Zhang, P., Zhao, J., Wang, B., Du, J., Lu, Y., Chen, J., and Ding, J.** (2006). The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein Sci.*
- Zhang, R., Calixto, C.P.G., Marquez, Y., Venhuizen, P., Tzioutziou, N.A., Guo, W., Spensley, M., Entizne, J.C., Lewandowska, D., Have, S. Ten, et al.** (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073.
- Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.H.** (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat. Cell Biol.*
- Zhou, Z., and Fu, X.-D.** (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122, 191–207.
- Zhou, H.L., Luo, G., Wise, J.A., and Lou, H.** (2014). Regulation of alternative splicing by local histone modifications: Potential roles for RNA-guided mechanisms. *Nucleic Acids Res.*
- Zhu, D., Mao, F., Tian, Y., Lin, X., Gu, L., Gu, H., Qu, L. jia, Wu, Y., and Wu, Z.** (2020). The Features and Regulation of Co-transcriptional Splicing in Arabidopsis. *Mol. Plant* 13, 278–294.
- Zhu, J., Liu, M., Liu, X., and Dong, Z.** (2018). RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis. *Nat. Plants.*



## Acknowledgments

I wish to express my sincere appreciation to Prof. Dr. Klaus D. Grasser. I am grateful for the consistent support and guidance during the running of this project, and in particular for the opportunity to conduct my PhD in his work group that became a second family for me.

I would like to thank Prof. Dr. Thomas Dresselhaus for leading my examination committee as chairperson and for the nice atmosphere in the department, I will always remember the Christmas duck and the fun playing Wichteln.

My sincere gratitude to my mentoring team: Prof. Dr. Jan Medenbach and Pr. Dr. Andreas Wachter for having always time for me providing guidance and feedback throughout this project. In addition I would like to thank Prof. Dr. Jan Medenbach for correcting my thesis and I am extremely grateful for our friendly chats at the end of our meetings and the encouragement in my project. Besides my advisors, I would like to thank Prof. Dr. Joachim Griesenbeck for the evaluation of my work as 3<sup>rd</sup> examiner.

Special thanks for the ChIPseq and bioinformatics analysis to Dr. Thomas Strepfl, Uwe Schwartz and especially to Simon Obermayer: sorry to be always the first Hacker of our cloud.

Also I thank Vroni, Sabine and Kinga for being competent and friendly in every bureaucratic situation.

Further, I would like to thank the current and former students of AG Grasser, and employees, in particular Antje and Michael. And the all department for the great atmosphere.

Thank you to Marion for being a close supervisor in my project willing for suggestions and persistent help in every moment.

From the bottom of my heart I would like to say big thank you to “the old KDGs” for being a second family for me. Without this wonderful group I would not have been able to complete this research, and made it through my PhD degree. For their energy, understanding and help throughout my project, especially to the grumpy “H”Ans, with whom I started this PhD experience and worked side by side until the really end, Wojtek for his precious advices widening my research from various perspectives, he taught me a lot during this years and we became “ChIP on ChIP” partner, Philipp for the insightful suggestions and the pretty comments almost every morning that made my days, Brian for our constructive discussion on “how to make a proper coffee” and Alex for being a great bench neighbor. I wish to remark also “the new promising KDGs”, Amelie, Valentin, Henna, Hanna and Simon.

To conclude, I wish to thank Marco, my husband. For being with me in every up and down moment of this experience with unconditional support. For listening to every presentation and pretending to understand. For his endless patience in every weak moment and for encouraging me to make it to the end. I cannot forget little Leo, my stronger supporter with his smiles and big eyes and my parents for always being the fun club of every decision in my life.