Doctoral Thesis

Histone modifications and transcript processing during Arabidopsis development

Author(s): Mahrez, Walid

Publication Date: 2013

Permanent Link: https://doi.org/10.3929/ethz-a-009970157

Rights / License: In Copyright - Non-Commercial Use Permitted
HISTONE MODIFICATIONS AND TRANSCRIPT PROCESSING DURING ARABIDOPSIS DEVELOPMENT

A dissertation submitted to ETH ZURICH for the degree of Doctor of Sciences

presented by

Walid Mahrez

MSc of Science Molecular Biology and Biochemistry (MAICh)

Born April 24, 1980

Citizen of Algeria

Accepted on the recommendation of

Prof. Dr. Whillem Gruissem, examiner

Prof. Dr. Ernst Hafen, co-examiner

Prof. Dr. Lars Hennig, co-examiner

2013
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABBREVIATIONS</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ZUSAMMENFASSUNG</td>
<td>4</td>
</tr>
</tbody>
</table>

## 1. INTRODUCTION

1.1 DNA packaging into chromatin 6
1.2 Histone modifications and histone code 8
1.2.1 Histone acetylation 8
1.2.2 Histone methylation 10
1.3. Transcription and transcript processing 12
1.3.1 Transcript capping and polyadenylation 12
1.3.2 Splicing 13
1.3.3 Alternative splicing 14
1.4. Flowering time control in the model plant Arabidopsis 14
1.4.1 Activation of FLC is associated with delayed flowering 16
1.4.2 Repression of FLC expression is associated with earlier flowering 16
1.5 Flowering as a agronomical trait 17
1.6 Goals of the thesis 18

## 2. MATERIAL AND METHODS

2.1 Materials 19
2.1.1 Mutants and transgenic lines of plants 19
2.1.3 Bacteria strains and plasmids 20
2.1.4 Enzymes 21
2.1.5 Antibodies 21
2.2 Methods 22
2.2.1 Growth conditions 22
2.2.2 Molecular methods 22
2.2.2.1 Methods for DNA 22
2.2.2.1.1 Plant DNA extraction with Edwards buffer 23
2.2.2.1.2 Molecular cloning 23
2.2.2.1.3 Identification of EMS-induced mutations in Arabidopsis by Illumina Deep Sequencing 23
2.2.2.1.4 Validation of the single nucleotide polymorphism in cääö 23
2.2.2.2 Methods for RNA 24
2.2.2.2.1 RNA Extraction and RT-PCR 24
2.2.3 Methods for proteins 24
2.2.3.1. Protein expression and purification from bacteria 24
2.2.3.2. Purification of 6x His-tagged proteins with TALON metal affinity resin 24
2.2.3.3. Purification of GST-tagged proteins with affinity chromatography 25
2.2.3.4 Histone extraction from plants identification by Tandem Mass Spectrometry 25
2.2.3.5 Histone fractionation by RP-HPLC and post translational modification
2.2.3.6 Mass Spectrometry database analysis
2.2.3.7 Protein separation on SDS-PAGE and Western-blotting
2.2.3.8 Dot blotting
2.2.4 Methods for chromatin
2.2.4.1 Crosslinked Chromatin Immunoprecipitation assays
2.2.4.2 Native ChIP-seq
2.2.5 Histone Acetyl Transferase activity Assay
2.2.6 Immunostaining
2.2.7 Sequence alignment and phylogenetic analysis

3. RESULTS
3.1 The BRR2a splicing protein is involved in Arabidopsis flowering time
Summary
Introduction
Materials and Methods
Results
Discussion
Acknowledgments
Supplement

3.2 H3K36ac is an evolutionary conserved active histone mark in plants
Summary
Introduction
Materials and Methods
Results
Discussion
Acknowledgments
Supplement

3.3 H3K23me is an evolutionary conserved histone modification associated with
gene silencing in Arabidopsis
Summary
Introduction
Materials and Methods
Results
Discussion
Acknowledgments
Supplement

4. Concluding discussion
REFERENCES
ACKNOWLEDGEMENTS
CURRICULUM VITAE
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>AGAMOUS</td>
</tr>
<tr>
<td>Brr2p</td>
<td>Bad response to refrigeration 2 protein</td>
</tr>
<tr>
<td>cäö</td>
<td>chrottäpösche</td>
</tr>
<tr>
<td>CO</td>
<td>CONSTANS</td>
</tr>
<tr>
<td>Col</td>
<td>Columbia</td>
</tr>
<tr>
<td>CRY</td>
<td>CRYPTOCHROME</td>
</tr>
<tr>
<td>DAPI</td>
<td>4p-6-diamidino-2-phenyl indole</td>
</tr>
<tr>
<td>dCAPS</td>
<td>Derived cleaved amplified polymorphic sequences</td>
</tr>
<tr>
<td>EOD</td>
<td>End of day</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FLC</td>
<td>FLOWERING LOCUS C</td>
</tr>
<tr>
<td>FT</td>
<td>FLOWERING LOCUS T</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GCN5</td>
<td>General control non-derepressible protein 5</td>
</tr>
<tr>
<td>HATs</td>
<td>histone acetyl transferases</td>
</tr>
<tr>
<td>HDACs</td>
<td>histone deacetylases</td>
</tr>
<tr>
<td>HD19</td>
<td>Histone deacetylase 19</td>
</tr>
<tr>
<td>HKMT</td>
<td>Histone lysin methyltransferase</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone H3 trimethylated at lysine 27</td>
</tr>
<tr>
<td>H3K23ac</td>
<td>Histone H3 acetylated at lysine 36</td>
</tr>
<tr>
<td>H3K23me</td>
<td>Histone H3 monomethylated at lysine 23</td>
</tr>
<tr>
<td>Jmj</td>
<td>jumonji domain</td>
</tr>
<tr>
<td>KYP</td>
<td>KRYPTONITE</td>
</tr>
<tr>
<td>LD</td>
<td>Long Day</td>
</tr>
<tr>
<td>Ler</td>
<td>Landsberg erecta</td>
</tr>
<tr>
<td>MAF</td>
<td>MADS AFFECTING FLOWERING gene family</td>
</tr>
<tr>
<td>PolII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>PP2A</td>
<td>Protein phosphatase 2 A</td>
</tr>
<tr>
<td>PRC</td>
<td>Polycomb repressive complex</td>
</tr>
<tr>
<td>PTM</td>
<td>Post Translational Modifications</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA binding proteins</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>snRNP</td>
<td>small nuclear ribonucleo-protein</td>
</tr>
<tr>
<td>SD</td>
<td>Short Day</td>
</tr>
<tr>
<td>SEP3</td>
<td>SEPALATA 3</td>
</tr>
<tr>
<td>SET</td>
<td>Su(var) 3-9, Enhancer of zeste and Trithorax</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SOC1</td>
<td>SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1</td>
</tr>
<tr>
<td>Ta</td>
<td>Transposon</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription Start Site</td>
</tr>
<tr>
<td>TRX</td>
<td>Trithorax</td>
</tr>
<tr>
<td>TTS</td>
<td>Transcription Termination Site</td>
</tr>
</tbody>
</table>
ABSTRACT

The switch from the vegetative to the reproductive phase, known as the flowering phase transition, is one of the most important developmental steps for flowering plants, and it constitutes a very important agronomical trait.

Because of the impact it has on crop yield, molecular mechanisms underlying the control of flowering time in plants have received much attention. Most of this research is performed in model species such as *Arabidopsis thaliana* and many examples of successful translation of this knowledge into crops exist.

Several pathways control the flowering time in Arabidopsis through transcriptional and posttranscriptional gene regulation, including transcript processing, chromatin modifications and chromatin remodeling. With the aim to better understand the molecular mechanisms of flowering time control, a mutant screen for accelerated flowering of Arabidopsis was previously performed in our laboratory. Here, one of the early flowering mutants was used for further investigations. Using Illumina deep sequencing, I identified the causal mutation as a transition from G to A in the *BRR2a* gene and further confirmed this finding by Sanger sequencing and dCAPS marker analysis as well as by an allelism test that used a T-DNA insertion allele. The BRR2a protein is a key component of the spliceosome complex; it is highly conserved throughout the eukaryotes, and ubiquitously expressed in all tested tissues in Arabidopsis. Using molecular tools, I identified the highly structured introns of *FLC* and some additional flowering time genes to be major targets of BRR2a. Mutant plant had altered transcript processing efficiency for *FLC*, and the reduced level of *FLC* mRNA could explain the observed increased transcription of *FT* and *SOC1* resulting in an early flowering in both short and long days. Taken together, a new RNA-binding and processing factor was identified to be involved in the Arabidopsis flowering transition at the posttranscriptional level.

To better understand the mechanism of gene expression in plants at the transcriptional level, I searched for additional uncharacterized histone posttranslational modifications that may affect the gene’s accessibility to RNA Polymerase II for transcription. In this work, I had used histone extracts from *Brassica oleracea* (cauliflower) to study the histone code in Brassicaceae.
Many histone marks known in Arabidopsis were found. In addition, I identified several additional modifications not characterize in plants before. Here, I focused on the detailed study of two such newly described histone modifications - H3K36ac and H3K23me1. These two marks were found to be evolutionary conserved modifications in plants and animals. Using ChIP-Seq, H3K36ac was found to be enriched at the 5’ end of actively transcribed genes. Independently of gene length, this modification covers about 500 bp immediately downstream of the transcriptional start site where it overlaps with the H2A.Z histone variant and with H3K4me3. For H3K36ac homeostasis the histone acetyl transferase GCN5 and the histone deacetylase HDAC19 are needed. On the other hand, H3K23me1 was identified as a heterochromatin localized mark, enriched on repetitive DNA sequences and transposons as well as on silent Polycomb target genes. I found the Kryptonite histone methyl transferase to be required for full H3K23me1 levels. In addition, a functional connection between H3K23me1 and DNA methylation on transposable elements was identified.

In summary, this work contributed to a better understanding of the mechanisms that regulate gene expression during plant development by the identification of a new Arabidopsis flowering time gene and two novel components of the plant histone code.
ZUSAMMENFASSUNG

Der Übergang von der vegetativen zur reproduktiven Phase, bekannt als die Blütephase-transition, ist einer der wichtigsten Entwicklungsschritte für Blütenpflanzen, welcher auch für die Agrarwirtschaft von großer Bedeutung ist. Wegen dem grossen Einfluss auf den Ernteertrag wird molekulare Mechanismen, die die Blütezeit steuern, intensiv studiert. Der grösste Teil dieser Forschung wird in Modellorganismen wie Arabidopsis thaliana durchgeführt, und es gibt viele Beispiele, wie dieses Wissen erfolgreich auf Kulturpflanzen angewandt werden konnte. Mehrere Signalwege steuern die Blütezeit in Arabidopsis durch transkriptionelle und posttranskriptionelle Genregulation, dazu gehören Transkriptmodifikationen, Chromatinmodifikationen und Chromatinrestrukturierung.


Die Mutante wies eine reduzierte Transkriptprozessierungseffizienz für FLC auf, und die reduzierte Menge an FLC mRNA kann die beobachtete erhöhte Transkription von FT und SOC1 erklären, die zu verfrühtem Blühen in Kurz- und Langtag führte. Zusammenfassend wurde ein neuer RNA-binde/-prozessierungs faktor identifiziert, welcher am Blühübergang auf posttranskriptioneller Ebene beteiligt ist.


Zusammenfassend kann man sagen, dass diese Arbeit durch die Identifikation eines neuen Blühzeitpunktkontrollgens und zweier Komponenten des pflanzlichen Histoncodes zum besseren Verständnis der Mechanismen beitragen, die die Genexpression während der Pflanzenentwicklung regulieren.
1. INTRODUCTION

Development of multicellular organisms depends on the strictly regulated realization of alternative gene expression patterns. In fact, expression of so-called marker genes is often taken as indication of cell-type identity. Failure to control gene expression often has strong negative consequences on phenotype and in humans can lead to diseases such as cancer (Chi et al, 2010). Gene expression is controlled at multiple levels allowing fine-tuned control as well as robustness. Major control levels are at (i) DNA packaging into chromatin, (ii) transcription, (iii) mRNA processing, (iv) mRNA stability, (v) translation and (vi) protein stability. Because DNA packaging into chromatin and mRNA processing are most relevant in the context of this thesis, they will be discussed in more detail in the next two sections. In a third section, the transition to flowering, which is a major developmental transition in plants and of relevance to this thesis, will be introduced.

1.1 DNA packaging into chromatin

Chromatin is the structure of DNA, protein and RNA that nuclear DNA in eukaryotes exists in and forms the substrate transcription as the first step of gene expression has to deal with. The basic unit of chromatin is the nucleosome, constituted of 147 bp of double stranded super helical DNA wrapped around an octamer formed by two copies each of four basic proteins called histones (H2A, H2B, H3, H4) (Fig.1.1) (Richmond and Davey, 2003, Luger et al, 1997) Individual nucleosomes can be linked by additional histones such as H1.
Figure 1. Chromatin structure and histone modifications

(A) Face view of the nucleosome (Protein Data Bank code 1AOI; using PyMOL software), constituted of a double stranded DNA backbones (brown) wrapped around histone proteins forming an octamer a pair of (four different histones H2A (orange/light pink), H2B (purple/grey), H3 (light Blue/yellow), H4 (green/dark pink) are displayed with different colors); (B) Top view of the nucleosome; the amino terminal side chains of histones are displayed and marked by asterisks; (C) Amino acid composition of the core nucleosome histone proteins, and examples of posttranslational modifications on specific residues (ac, Acetylation; Ph, Phosphorylation; me, Methylation; cit, Citrullination; Pr, Propionylation), the numbers below the amino acids correspond to the position of the residues at the amino or carboxy terminal site of the histone proteins.

Histones are small proteins of (11 to 21.5 kDa), which are highly conserved from yeast to humans (Mcghee and Felsenfeld, 1980). Histones are composed of a globular carboxy-terminal domain and flexible amino-terminal “tails” that protrude from the nucleosome surface and can act to recruit protein complexes, which often harbor enzymatic activities. The basic nature of histones can attract the negatively charged DNA, tightly bind to it and make it fit into the nucleus (Richmond and Davey, 2003)

Chromatin is unequally distributed in interphase nuclei forming two main cytologically distinct regions upon staining of DNA: (i) weakly stained regions, which are defined as euchromatin, and (ii) strongly stained regions, which are defined as heterochromatin (Heitz, 1928).
At the molecular level, euchromatin was found to be gene-rich and highly accessible for transcription, whereas heterochromatin was found to be gene-poor and instead rich in repetitive sequences (Elgin and Grewal, 2003).

Several protein complexes establish additional compaction or relaxation of local chromatin structures thus making specific DNA sequences either more or less accessible to transcription factors and RNA polymerases (Conaway, 2012). DNA accessibility in chromatin strongly affects cellular processes, that use DNA namely transcription, replication, recombination and DNA repair. Local chromatin properties and accessibility are controlled by two main groups of protein complexes: chromatin modifiers and ATP-dependent chromatin remodelers. Chromatin modifiers deposit or remove particular chemical entities on specific nucleotides in the DNA or on specific amino acid in histones. These modifications can lead to altered physicochemical properties of chromatin causing opening or compacting (Bannister and Kouzarides, 2011). Alternatively, these modifications can create landmarks for the recruitment of additional protein complexes. The ATP-dependent chromatin remodelers are a set of enzyme-containing complexes that use ATP hydrolysis to unwrap, slide or eject nucleosomes (Clapier and Cairns, 2009). Together these sets of complexes facilitate or prevent access of regulatory proteins to the DNA and thus facilitate or repress transcription.

1.2 Histone modifications and histone code

Both the globular core regions as well as the protruding amino-terminal tail of histones are extensively targeted by numerous posttranslational modifications (PTM). Over one hundred distinct modifications have already been described in the literature (Rando, 2012). These modifications are characterized by the deposition of small chemical groups (e.g. acetylation, phosphorylation, biotinylation and methylation) or larger peptides (e.g. ubiquitylation and sumoylation) by specific protein complexes, which are sometimes called “writers”. Once a modification is established, it constitutes a hallmark for the recognition by other proteins, which are sometimes called “readers”. The readers can recruit additional protein complexes, which establish additional histone marks in the local chromatin neighborhood or remove marks that are already established. Enzymes that can remove marks are sometimes called “erasers” (Chi et al, 2010). Writers, readers and erasers are all chromatin modifiers.

Research during the last decade or two has established that histone modifications form an important regulatory layer for processes such as gene expression, DNA replication and repair, chromosome condensation and segregation (Fullgrabe et al, 2011). The combinatorial nature of histone modifications was rationalized in the idea of a “histone code” over a decade ago (Strahl and Allis, 2000). The histone code, however, should be considered as a concept of thought rather than a description of biological reality, as results from high-resolution genome-wide mapping as well as high-sensitivity proteomic and genomic methods do not lend support to such a simplistic view of this issue (Rando, 2012).

The particular outcome of a histone modification depends largely on 1) Type of modifications, 2) the organism, in which the modification occurs, 3) the modified residue, 4) the degree of the modification, and 5) the chromatin context. Genome-wide analysis in different organisms established precise localization of several modifications at gene bodies or at promoters and regulatory elements, as well as the functional significance of co-localizing histone marks at these regions (Roudier et al, 2011, Kharchenko et al, 2011, Ernst and Kellis, 2011). Two of the most widely studied histone modifications are described below.

1.2.1 Histone acetylation

Histone acetylation is a dynamic, reversible chemical process that is highly conserved among eukaryotes. It involves the transfer of an acetyl group from acetyl-CoA to the ε-amino group of histone lysine residues.
(Fig. 1.2). The acetylation of a set of conserved lysine residues neutralizes the positive charge of histone tails and may decrease their affinity for negatively charged DNA (Hong et al., 1993). Once established, acetylated lysines can be recognized by histone code readers such as Bromo-domain proteins (Filippakopoulos and Knapp, 2012). These often promote the recruitment of additional ATP-dependent chromatin remodelers and chromatin modifier complexes, which establish a relaxed chromatin that facilitates the recruitment of RNA polymerases and gene expression (Bannister and Kouzarides, 2011).

As reported in the literature, histone lysines acetylation, which is often associated with active transcription, plays an important role in numerous developmental and biological processes in plants, namely regulation of cell cycle, flowering time, response to environmental conditions and hormone signals (Chen and Tian, 2007).

Amino-terminal lysine residues of histone H3 (K4, K9, K14, K18, K23, K27 and K36) and H4 (K5, K8, K12, K16, and K20) were reported to be acetylated in Arabidopsis as well as in other plants, yeast and animals (Berr et al., 2011).

Histone acetylation is dynamically regulated with a half-life of usually 2-3 minutes and rarely exceeding 30-40 minutes (Barth and Imhof, 2010). The homeostasis of the acetylation on histone lysines is maintained by histone acetyl transferases (HATs), which catalyze the histone acetylation, and histone deacetylases (HDACs), which remove the acetylation mark (Pandey et al., 2002). Histone acetylation frequently co-localizes with HATs and correlates with Pol II binding and gene expression (Tie et al., 2009, Horakova et al., 2010, Wang et al., 2009).

**Figure 1.2. Histone lysine modifications.**

The ε-amino group of histone lysine residue can be targeted by acetylation, mono-, di- or trimethylation. CH₃, methyl group; HAT, histone acetyltransferase; HDAC, histone deacetylases; HMT, histone methyl transferases; HDM, histone demethylase.
Chromatin immunoprecipitation coupled to microarrays or next generation sequencing (ChIP-chip and ChIP-Seq) were used for genome-wide profiling of histone acetylation in yeast, plants and animals revealing that H3K4ac, H3K9ac, H3K14ac, H3K18ac and H4K16ac showed strong peaks at the promoters of active genes. In addition, it was suggested that the histone modification patterns are highly informative in regard to the location of nucleosomes and state of associated genes along the chromosomes (Liu et al, 2005). Furthermore, the degree of acetylation and the acetylated residues may influence the transcriptional rate genome-wide (Guillemette et al, 2011, Pokholok et al, 2005, Liu et al, 2005).

1.2.2 Histone methylation

Unlike acetylation, histone lysine methylation does not alter the charge of the modified residues on histone tails. Lysine methylation can involve one (me1), two (me2) or three (me3) methyl groups per ε-amino group (Fig. 1.2). It is associated with both active and inactive genes and has different cytological localization depending on the methylated residue, the degree of methylation, the position on the gene body, the chromatin context as well as the species (Liu et al, 2010a). For example, in mice, H3K9me3 and H4K20me3 are abundant in condensed heterochromatin and correlate positively with gene silencing whereas mono- and dimethylated H3K9 are enriched in euchromatin (Robin et al, 2007).

In contrast, H3K9me3 and H4K20me3 are located in euchromatin and associate with gene expression in Arabidopsis (Charron et al, 2009). Furthermore, H3K9me1 and H3K9me2 localize to heterochromatin and mark silenced genes in Arabidopsis (Fransz et al, 2006). Histone methylation has important roles in many biological processes, including transcription and cell-cycle regulation, DNA damage and stress response, heterochromatin formation and X-chromosome inactivation, development and differentiation (Greer and Shi, 2012, Martin and Zhang, 2005).

Histone methylation can not only target lysines but also arginines and histidines (Greer and Shi, 2012). Methylation on histidines is least well understood. As mentioned above, lysines can be mono-, di- or trimethylated; and arginines can be mono- or dimethylated. With arginine dimethylation, both methyl groups can be on the same terminal nitrogen atom of the guanidinium group (asymmetric dimethylation) or one methyl group can be on each of the two terminal nitrogens (symmetric dimethylation). Using mass spectrometry and proteomic approaches, the histone H3 lysines K4, K9, K27, K36, K79 and H4 lysine K20, as well as H3 arginines R2, R8, R17, R26 and H4 arginine R3 were found to be methylated in yeast, ciliates, flies, plants, and mammals. However, distribution and biological outcome may differ among these organisms (Greer and Shi, 2012, Di Lorenzo and Bedford, 2011).

The homeostasis of the methylation on histone lysines and arginines is maintained by two distinctive enzymes: HKMTs (histone lysine methyl transferases) and PRMTs (protein arginine methyl transferases) catalyze the transfer of methyl groups from S-adenosylmethionine to the ε-aminogroup of lysine or to the terminal guanidino nitrogen atoms of arginines, respectively. Histone demethylases remove the methyl groups (Liu et al, 2010a, Bedford, 2007). Histone methylation on lysine is catalyzed by at least two enzymes families: SET domain-containing proteins (Rea et al, 2000), which add the methyl group on most lysines, and DOT1-like proteins, which methylate only H3K79 (Feng et al, 2002). Histone lysine methylation is removed by two families of histone demethylases: FAD-dependent monoamine oxidases and jumonji C (JmjC)-domain-containing iron-dependent dioxygenases (Chen et al, 2011, Spedaletti et al, 2008). Enzymes of the above mentioned groups are highly conserved from yeast to humans and have substrate specificities that often include both histone and non-histone proteins. The disruption of these enzymes often leads to developmental defects, diseases such as cancer and cognitive disorders and premature ageing (Greer and Shi, 2012).
Once established, methylated lysines constitute landmarks for recognition by the aromatic cages of proteins with methyl-binding domains (Taverna et al, 2007) such as PHD fingers, WD40 repeats, CW domains (Hoppmann et al, 2011), PWWP domains, ankyrin repeats (Collins et al, 2008), chromodomains, double chromodomains, chromobarsrels, Tudor domains, double or tandem Tudor domains and MBT repeats. These protein readers subsequently recruit additional protein complexes, to establish defined chromatin states (Greer and Shi, 2012). Genome-wide profiling using ChIP-chip or ChIP-seq in Arabidopsis (Roudier et al, 2011), Drosophila (Kharchenko et al, 2011) and in mammals (Ernst and Kellis, 2011) led to the definition of “chromatin states, CS”, distinguishable by their histone methylation patterns.

Histone methylation on particular lysine residues, including H3K9me2, H3K4me3, H3K36me2/3, and H2K27me3, has recently received much attention mostly due to their prominent roles in regulating gene expression and chromatin states as well as in controlling development. Some insights about H3K4me3 is given below.

H3K4 is a highly conserved residue, which is susceptible to both methylation and acetylation in a wide range of organisms including yeast, plants and animals (Guillemette et al, 2011). In Arabidopsis two-thirds of the genes contain at least one type of H3K4methylation. While H3K4me1 is abundant on the entire transcribed region, H3K4me2 and H3K4me3 accumulate predominantly in promoters and at 5' ends of genes (Zhang et al, 2009).

Furthermore, together with H3K36me3, H3K4me3 defines the chromatin state 1 (CS1), which groups active genes (Roudier et al, 2011). Similar to plants, H3K4me3 is highly enriched at the transcriptional start site (TSS) and has been correlated with transcriptionally competent or active genes in mammals. H3K4me3 promotes transcription through interaction with effector proteins including transcription factors to stimulate the recruitment of RNA polymerase II to the promoter of target genes (Lauberth et al, 2013).

In yeast, only one H3K4 methyltransferase (SET1; SET domain containing 1) was identified to establish all methylation states on H3K4. Multiple putative H3K4 methyltransferases homologous to SET1 have been identified in Arabidopsis and are believed to act in distinct complexes to establish the differently methylated states (Zhang et al, 2009). These enzymes belong to the Trithorax group proteins (TRX) (Thorstensen et al, 2011). They are highly evolutionary conserved proteins and were first identified in Drosophila (Ingham, 1998). They seem to play a role in methylating histones and remodeling chromatin for instance at homeobox genes to maintain them in an active state. As recently observed, they are also involved in the control of tumorigenesis and embryonic stem cell self-renewal, cell fate choice and proliferation, X-chromosome inactivation, apoptosis, cell cycle regulation, growth plasticity and regeneration as well as in stress responses (Schuettengruber et al, 2011).

The Arabidopsis genome encodes five TRX homologs and seven TRX related proteins (Thorstensen et al, 2011). ARABIDOPSIS TRITHORAX 1 (ATX1) is an active histone methyltransferase specific for histone H3K4 (Alvarez-Venegas et al, 2003). Its mutation leads to abnormal floral organ identity and slightly early flowering. It was reported that ATX1 trimethylates H3K4 at the promoter and the first exon of the FLC locus (Liu et al, 2010a) and thus controls plant flowering time (see below). H3K4me1 and H3K4me2 are removed in Arabidopsis by LDL1, LDL2 and FLD, homologues of human LSD1 and Drosophila SU(VAR)3-3 (Jiang et al, 2007).
1.3. Transcription and transcript processing

Transcription, i.e. the copying of genetic information of DNA into RNA, is carried out by RNA polymerases. Three types of RNA polymerases exist in eukaryotes: RNA polymerase I (Pol I), which transcribes rRNA genes; Pol II, which transcribes all protein coding genes as well as several small nuclear RNAs, and Pol III, which transcribes 5S rRNA, tRNAs and some other small RNAs.

Transcription of genes by Pol II occurs in three phases: initiation, elongation and termination, and involves many auxiliary proteins (Lewin et al, 2011). Initiation is a highly controlled process characterized by the recruitment of the general transcription factors (GTFs), which help in the recognition of the promoter sequences and thus the exact positioning of Pol II onto the DNA. GTF aid the opening of the double-stranded DNA and allow transcription to initiate. Subsequently, different elongating transcriptional factors contribute to the sliding of the polymerase along the DNA strand and to the elongation of the transcript synthesis. Furthermore, several cotranscriptional processes occur during elongation, including the capping of the messenger RNA at the 5’ end by a methylated guanine nucleotide as well as the removal of introns by the splicing process. The termination of transcription is aided by termination transcription factors and is marked by the cleavage and polyadenylation of the transcript. The C-terminal domain (CTD) of RNA polymerase Pol II plays an important regulatory function during the different phases of transcription. CTD has species-dependent number of highly conserved heptapeptide repeats of the sequence (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7). The distinct, dynamic phosphorylation of the three serines, one threonine and one tyrosine plays an important role in the loading of transcription factors and mRNA processing factors to Pol II along the transcription process (Kim et al, 2010, Chapman et al, 2008, Perales and Bentley, 2009).

The final transcript will be quality-controlled by RNA quality control pathways including the nonsense mediated decay (NMD) pathway (Brogna and Wen, 2009) and exported to the cytoplasm, where translation, storage and, ultimately, mRNA degradation occur. Most of the processing mechanisms on the nuclear pre-mRNAs at the 5’ ends (capping), 3’ ends (cleavage and polyadenylation) and internally (splicing) appear to be functionally coupled to transcription (Perales and Bentley, 2009) and mediated by nuclear RNA Binding Proteins (RBPs) (Bailey-Serres et al, 2009).

1.3.1 Transcript capping and polyadenylation

Transcript capping occurs on the 5’ ends of 25–50 bases long nascent transcripts, in a three-step process and in an RNA sequence-independent manner (Perales and Bentley, 2009). First, RNA triphosphatase removes the γ-phosphate of the first nucleotide, and then Guanylyl Methyl Phosphate (GMP) is added by RNA guanylyl transferase. Finally, the guanine is methylated at N7 by a methyl transferase. When transcription initiates, Pol II phosphorylated on the Ser5 residue at its CTD loads the capping enzymes on the mRNA and permits the capping. The relationship between capping enzymes and Pol II illustrates the communication between processing factors and the transcription machinery (Perales and Bentley, 2009).

3’ end processing of most mRNAs is a two-step reaction comprising endonucleolytic cleavage shortly after the AAUAAA sequence, followed by polyadenylation of the exposed 3’ OH (Perales and Bentley, 2009). Homologous multi-subunit complexes, including cleavage stimulation factor (CstF) and cleavage polyadenylation specificity factor (CPSF) in mammals and cleavage factor 1A (CF1A) and cleavage polyadenylation factor (CPF) in yeast, perform coupled cleavage and polyadenylation (Perales and Bentley, 2009). Furthermore, serine 2 phosphorylation on the CTD of Pol II may positively affect binding
of the 3’ end-processing factors and stimulate cleavage and polyadenylation of transcripts (Ahn et al., 2004).

1.3.2 Splicing
Splicing is carried out by one of the most complex cellular machineries, the spliceosome, a large protein–RNA complex of more than 100 protein subunits and five small nuclear ribonucleic particles (snRNPs). The splicing reaction occurs via the recognition of highly conserved short sequences on the pre-mRNA namely the 5’ splice site, the branch-point, poly-pyrimidine tract and the 3’ splice site. In higher eukaryotes, these sequences can be poorly conserved; thus, additional auxiliary sequences within the pre-mRNA can direct splice site selection. These sequences either enhance or repress splicing and can be located in exons or introns. Folding of pre-mRNAs to secondary structures also affects splice site selection (Warf and Berglund, 2010) (Fig. 1.3).

Splicing of pre-mRNAs can be functionally divided into several steps, including spliceosome assembly and activation, catalysis and disassembly of the spliceosomal machinery (Will and Luhrmann, 2011). The spliceosome is first assembled by the interaction of the U1 snRNP with the 5’ splice site, and then the U2 snRNP recognizes and binds to the intron branch point sequence forming the pre-spliceosome. Spliceosome assembly is completed by the subsequent association of the U4/U6 and U5 snRNPs in the form of a U4/U6·U5 tri-snRNP complex. Activation of the spliceosome occurs by the disruption of U4 and U6 base-pairing and the formation of U2 and U6 base-pairing. Concomitantly with these events, base-pairing of U1 with the 5’ splice site is exchanged for base-pairing between U6 and the 5’ splice site. After these rearrangements, the U1 and U4 snRNPs are released from the spliceosome prior to catalysis. The conserved loop 1 of U5 RNA makes contact with exonic sequences at the 5’ and 3’ splice sites while the splicing reaction proceeds. Once the splicing reaction is completed, the spliced mRNA is released and the spliceosome is disassembled in an ATP-dependent manner (Bottner et al., 2005).

Figure 1.3. Mechanisms of Pre-mRNA Splicing
Schematic representation of the different splicing steps catalyzed by the spliceosomal proteins, including eight evolutionarily conserved DExD/H-type RNA-dependent ATPases/helicases; 5’ss, splice site, BP, Branching point; U1,2,3,4,6, snRNP [modified after (Will and Luhrmann, 2011)]

1.3.3 Alternative splicing

Alternative splicing (AS) is an important posttranscriptional regulatory mechanism in metazoans and plants by which multiple forms of mature mRNAs are generated from a single pre-mRNA. AS plays critical roles in differentiation, development and disease and is a major source for protein diversity in higher eukaryotes. AS events in plants were initially identified to affect genes involved in splicing, transcription, flowering regulation, disease resistance, metabolic activities and many other physiological processes. The most recent genome-wide analysis of AS events in Arabidopsis, which were based on RNA-seq, had revealed that at least \( \sim 42\% \) of transcribed pre-mature mRNAs were alternatively spliced (Wang and Brendel, 2006, Filichkin et al, 2010).

One of the first discovered and best-studied alternative splicing events in plants affects the flowering time control gene \( FCA \) (Macknight et al, 2002). \( FCA \) has 20 introns and encodes a protein containing two RNA-recognition motifs and a WW protein interaction domain. The \( FCA \) transcript is alternatively processed at two positions, resulting in four transcripts: \( \alpha \), \( \beta \), \( \gamma \) and \( \delta \) (Macknight et al, 2002).

Differential processing of intron 3 yields three different transcripts: Transcript \( \alpha \) retains intron 3; transcript \( \beta \) has premature cleavage and polyadenylation within intron 3; transcript \( \gamma \) has intron 3 spliced out. Alternative splicing on intron 13 forms transcript \( \delta \). The alternative forms of \( FCA \) are present in different ratios in different tissues of Arabidopsis, and only the transcript \( \gamma \) encodes a functional FCA protein that represses \( FLC \) and favors flowering (Macknight et al, 2002). Because splicing occurs co-transcriptionally, transcription elongation rates can strongly affect the selection of alternative splice sites and transcript processing efficiency in general (Kornblihtt et al, 2004). Local chromatin properties strongly affect not only transcription initiation but also elongation (Joshi and Struhl, 2005), and not surprisingly chromatin properties can control splicing. It is thought that transcription elongation has to be slowed down at genes with many introns to allow sufficient time for correct and complete splicing (de la Mata et al, 2011). In plants, the process that has most contributed to our current understanding of how chromatin and RNA processing affect gene expression and development is the transition to flowering (Eckardt, 2002).

1.4. Flowering time control in the model plant Arabidopsis

The switch from vegetative growth to reproduction is an important developmental step for plants. The exact timing to flower is influenced by numerous factors including (i) endogenous cues such as hormones and plant age and (ii) environmental stimuli such as light quality, photoperiod and temperature.

Four main genetic pathways, the photoperiod, vernalization, gibberellic acid (GA) and autonomous pathway, are established to control the flowering transition in Arabidopsis (Fig. 1.4). The photoperiod pathway promotes flowering in the favorable long-day photoperiod conditions of spring (Andres and Coupland, 2012). The vernalization pathway establishes a biannual lifestyle with flowering occurring only after completion of one winter (Lin et al, 2005). The gibberellic acid and autonomous pathways are thought to promote flowering upon favorable internal metabolic conditions (Wilson et al, 1992, Simpson, 2004, Wahl et al, 2013).

The different flowering pathways converge to control a common set of genes called floral integrators including in particular \textit{SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1} (SOC1) and \textit{FLOWERING LOCUS T} (FT), which both are potent activators of flowering (Srikanth and Schmid, 2011).
The autonomous and vernalization pathways converge to repress *FLC* – the major repressor of the floral integrators (Srikanth and Schmid, 2011).

Figure 1.4 Pathways controlling flowering time in *Arabidopsis*.

The autonomous pathway includes chromatin modifiers such as FVE, FLD and HDA6 and RNA binding proteins such as FY, LD, FPA, FCA and FLK to repress *FLC*. The vernalization pathway includes chromatin modifiers such as VIN3, VRN1 and VRN2 to repress *FLC*. The FRI complex together with the PAF1 complex promotes *FLC* expression. *FLC* represses the floral integrators *SOC1* and *FT*. The photoperiod and gibberellin pathways promote expression of the integrators leading to the induction of floral-meristem-identity genes such as *LEAFY* and *AP1* and thus to flowering. Lines with arrow heads indicate activation of gene expression and lines with bars, indicate gene repression.
1.4.1 Activation of FLC is associated with delayed flowering

FLC encodes a MADS box DNA binding protein, the main repressor element in the control of flowering time (Michaels and Amasino, 1999). Late flowering ecotypes of Arabidopsis carry active alleles of FLC and FRI, whereas early-flowering ecotypes contain a recessive allele of either one or both genes (Johanson et al, 2000). FRI-mediated FLC activation requires different complexes that modify FLC chromatin. The RNA Polymerase II (Pol II) Associated Factor 1 complex (PAF1c) was identified as an evolutionary conserved protein complex from yeast to humans. PAF1c is devoid of any enzymatic activities (Oh et al, 2004), and it acts as a scaffold on which many chromatin modifying complexes, as well as RNA Pol II bind and establish active histone marks (H3K4me2/3 and H3K36me2/3) on the FLC chromatin, thus enhancing its expression (Yu and Michaels, 2010).

This protein complex was described to function together with the FRI activating complex (He, 2012), which contains at least four components FRL1, FES1, SUF4 and FLX. Together with FRI, they constitute a transcription activator complex (FRIc) that subsequently recognizes the proximal promoter of FLC and recruits additional histone modifiers and chromatin remodelers to activate FLC expression (Choi et al, 2011).

The recruitment of the COMPASS H3K4 methyltransferase complex and the EFS H3K36 methyltransferase onto FLC TSS and gene body, respectively, leads to the trimethylation of their target residues (H3K4 and H3K36, respectively) leading to an establishment of active chromatin states favoring high FLC expression (Jiang et al, 2011, Xu et al, 2008). In addition to histone modifications, the deposition of active histone variants e.g. H2A.Z at the 5' end of FLC by the ATP dependent chromatin remodeler complex SWR1c (Deal et al, 2007) favors the transcriptional activation of the flowering repressor FLC leading to delayed flowering (He, 2012).

1.4.2 Repression of FLC expression is associated with earlier flowering

The autonomous pathway genes act as FLC repressors (He and Amasino, 2005), and they belong to two main subfamilies: (1) chromatin modifiers such as FLOWERING LOCUS D (FLD); (He et al, 2003), FVE (Ausin et al, 2004), RELATIVE OF EARLY FLOWERING6 (Noh et al, 2004), Polycomb (PcG) components (Jiang et al, 2008), HISTONE DEACETYLASE 6 (HDA6) (Yu et al, 2011) and (2) putative RNA binding proteins (RBP) such as FCA (Simpson et al, 2003), FPA (Schomburg et al, 2001), FY (Simpson et al, 2003), and FLOWERING LOCUS K (Lim et al, 2004), LD (luminodependance) (Lee et al, 1994).

Members of the last subfamily carry specific motifs to recognize and bind RNA such as RNA recognition motifs (RRM) and K-Homology domains (KH). They bind and affect processing of FLC transcripts in a still poorly understood way (Lorkovic, 2009).

FLD in complex with HDA6 and FVE exerts a repressive effect on the FLC locus by specifically demethylating H3K4me3, and deacetylating acetylated lysines on the promoter and around the TSS of FLC (Gu et al, 2011), allowing repressive marks to be deposited such as H3K27me3 by Polycomb repressive complex 2 (PRC2) (Jiang et al, 2008). H3K27me3 then contributes to long-term repression of FLC and initiation of flowering (He, 2012).
Winter annuals plants need an extended cold exposure (vernalization) prior to flowering (Zografos and Sung, 2012). At the molecular level, this is achieved by a long-term, stable silencing of \( FLC \) that requires both long non-coding RNAs (lncRNAs) and Polycomb group protein complexes (Jiang et al, 2008, Liu et al, 2010b).

For stable \( FLC \) silencing after vernalization, a PRC2-like repressive complex incorporates PHD domain-containing proteins such as VERNALIZATION INSENSITIVE 3 (VIN3) and VRN5 to form a complex named PHD-PRC2 (De Lucia et al, 2008). PHD-PRC2 formation is induced upon vernalization and required to deposit H3K27me3 across the entire \( FLC \) locus needed for stable silencing and flowering induction (De Lucia et al, 2008). In addition to the coding \( FLC \) transcript, the \( FLC \) locus can produce a sense non-coding transcript from within intron 1 (COLDAIR, COLD ASSISTED INTRONIC NONCODING RNA (Heo and Sung, 2011)) and two different antisense non-coding transcripts [COOLAIR, COLD INDUCED LONG ANTISENSE INTRAGENIC RNA (Swiezewski et al, 2009)]. The two COOLAIR transcripts have alternative 3' end polyadenylation sites, which are formed by FCA, FPA and the RNA 3'-end processing factors FY, CstF64, and CstF77 (Liu et al, 2010c). By a poorly understood mechanisms, this co-transcriptional antisense RNA 3'-end processing leads to \( FLC \) silencing (He, 2012).

In addition to the above described pathways and factors, additional proteins and even pathways are continuously being identified as key players in the regulation of flowering time (Ietswaart et al, 2012, Wahl et al, 2013). Some floral regulators can also function in more than one flowering pathway such as the histone binding protein MSI1, which was found to be required for normal timing of flowering but could not be assigned to a single genetic pathway (Bouveret et al, 2006). Transgenic plants expressing a TAP-tagged version of \( MSI1 \) in an \( msi1 \) mutant background (\( msi1\)-tap1) were used before as tool to study the role of \( MSI1 \) in flowering control (Bouveret et al, 2006, Alexandre et al, 2009, Exner, 2008).

With the aim to better understand the molecular mechanism of flowering time control, a mutant screen for suppressors of the late flowering phenotype of \( msi1\)-tap1 plants was performed (Exner, 2008). Eleven families with a shortened vegetative phase of \( msi1\)-tap1 were selected for further studies. Two mutated genes were previously identified by mapping as \( LHP1 \) (Exner et al, 2009) and \( CRY1 \) (Exner et al, 2010). \( LHP1 \) is a plant Polycomb group protein that is required for \( FT \) silencing, and increased expression of \( FT \) causes early flowering in \( lhp1 \) mutants (Kotake et al, 2003). \( CRY1 \) is a blue-light photoreceptor, which is also involved in \( FT \) activation (Exner et al, 2010). Nine other mutant families remained to be characterized.

1.5 Flowering as a agronomical trait

The transition to flowering is a major change in plant development that does not only affect morphology but also metabolism and other aspects of physiology. Not surprising has flowering time great agronomical importance and is the target of both conventional breeding and modern biotechnology applications (Jung and Muller, 2009). Accelerated flowering is, for instance, desired to accelerate breeding programs for perennials such as fruit trees. In contrast, delayed flowering is desired for corps where vegetative parts are used and is an important breeding aim for sugar beet and forage grasses (Jung and Muller, 2009). Finally, climate change is expected to shift growing seasons in many areas and adjustments of flowering time to new yield optima will be needed. Therefore, basic research on flowering time control receives considerable attention by plant breeders for translational research applications.

Together, transition to flowering is not only an important model process for plant developmental biology and an important agronomical trait but is also a powerful tool to decipher mechanisms of transcriptional regulation.
Goals of the thesis

In order to gain a better understanding of the mechanism of flowering time control in model organism Arabidopsis thaliana, the objective of this study was to search for novel uncharacterized regulators of flowering time. This work could benefit from a mutant screen for early flowering plants performed previously in the laboratory (Exner, 2008). Two mutants form this screen were characterized before and had given valuable biological insights (Exner et al, 2009, 2010) but several non-characterized mutants remained. The aim was to select a mutant with a consistent and heritable phenotype. I will establish a mapping population and map the causal mutation by sequencing (Austin et al, 2011). The identified mutation will then be confirmed, and subsequent functional studies will attempt to uncover the molecular mechanisms behind the early flowering phenotype.

On the other hand, it was planned to expand our knowledge about histone posttranslational modifications as important regulators of flowering time and other developmental processes in plants. In yeast and animals, new histone posttranslational modifications are continuously being discovered, and it is likely that also in plants not all histone posttranslational modifications are known. In this study, I aim to search for additional uncharacterized histone marks in plants using accurate and high precision mass spectrometry technology. The identification of any novel histone mark will be followed by a set of molecular and biochemical studies to better understand its localization at the chromatin and gene level and its potential role in the regulation of gene expression in plants.

Altogether, I expect that this work will contribute to a better understanding of the mechanism of gene regulation during plant development.
2.1 Materials

2.1.1 Mutants and transgenic lines of plants

List of mutants and transgenic lines used in this study are described in Table 2.1.

All the transgenic lines listed below are in the Col-0 background unless otherwise stated.

Table 2.1 List of mutants and transgenic lines used in this study

<table>
<thead>
<tr>
<th>Gene ID/Construct/Alias</th>
<th>Allele</th>
<th>Mutagen</th>
<th>Original seed source ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT4G38130, 35S:HDA19-GFP</td>
<td></td>
<td>Transgene (Zhou et al, 2005)</td>
<td></td>
</tr>
<tr>
<td>HDA19-RNAi1</td>
<td></td>
<td>Transgene (Zhou et al, 2005)</td>
<td></td>
</tr>
<tr>
<td>AT3G54610, GCN5</td>
<td>gcn5-1</td>
<td>TDI*</td>
<td>Benhamed et al, 2006</td>
</tr>
<tr>
<td>AT5G63110, HDA6, RTS</td>
<td>hda6-1</td>
<td>TDI</td>
<td>Aufsatz, et al. 2002</td>
</tr>
<tr>
<td>AT5G63110, 35S:Flag-HDA6</td>
<td></td>
<td>Transgene (Earley, et al. 2006)</td>
<td></td>
</tr>
<tr>
<td>AT1G77300, SDG8, EFS</td>
<td>sdg8-2</td>
<td>TDI</td>
<td>SALK_026442</td>
</tr>
<tr>
<td>AT5G13960, KYP</td>
<td>Kyp-lh1</td>
<td>TDI</td>
<td>SALK_105816</td>
</tr>
<tr>
<td>0.3 185 cão/MSI1-TAP × Col</td>
<td>cão</td>
<td>EMS*</td>
<td>VE942 (Exner, et al. 2008)</td>
</tr>
<tr>
<td>0.3 185 cão/MSI1-TAP × Ler</td>
<td>cão</td>
<td>EMS</td>
<td>VE915.1 (Exner, et al. 2008)</td>
</tr>
<tr>
<td>0.3 185 cão/MSI1-TAP</td>
<td>cão,</td>
<td>EMS</td>
<td>VE355-1 (Exner, et al. 2008)</td>
</tr>
<tr>
<td></td>
<td>msl1-tap1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G00650, FRI in Col</td>
<td></td>
<td>Transgene (Bouveret, et al. 2006)</td>
<td></td>
</tr>
<tr>
<td>AT5G10140, FLC</td>
<td>flc6-1</td>
<td>TDI</td>
<td>(Bouveret, et al. 2006)</td>
</tr>
<tr>
<td>AT5G44800, EMB1507</td>
<td>emb1507-4</td>
<td>TDI</td>
<td>NASC ID: N16092</td>
</tr>
</tbody>
</table>

* TDI: T-DNA Insertion, EMS: Ethyl Methane Sulfonate, NASC: Nottingham Arabidopsis Seed Stock Centre

2.1.2 Oligonucleotide primers

All primers used in this study were obtained from Sigma-Aldrich (Germany), and are listed in Tables 2.2 and Supplemental tables. Primers used for molecular cloning were designed using the primer 3 software at (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and checked for secondary structures and melting temperature with PerlPrimer software. To genotype T-DNA insertion mutants, primers were designed using online software (http://signal.salk.edu) (Alonso et al, 2003).

For the design of dCAPS primers the dCAPS Finder 2.0 software (Neff et al, 2002) (http://helix.wustl.edu/dcaps/dcaps.html) was used. RT-PCR primers were designed through the Universal Probe Library Assay Design Center (Roche Applied Science).
Table 2.2 List of gene-specific primers used for genotyping of mutants

<table>
<thead>
<tr>
<th>Mutant allele</th>
<th>Primer ID/Forward primer sequence</th>
<th>Primer ID/Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hda6-1</td>
<td>LH0587: GATTCTGAGTGAGAGACGGAG</td>
<td>LH0588: AGCCATACGGATCCGCTGAGG</td>
</tr>
<tr>
<td></td>
<td>LH1366: GCTGGGGGTATTTATGAGAAG</td>
<td>LH1367: CACTGTCCAGTAAGCTGAGC</td>
</tr>
<tr>
<td>sdg8-2</td>
<td>LH0135: GATTCTAGGTTATAACGAGG</td>
<td>LH0136: GATGCCATGCAACATCTTCAC</td>
</tr>
<tr>
<td>msi1-1</td>
<td>LH0314: GCGAATTGAGAACAAGTATGCC</td>
<td>LH0571: CAATTCCAAGGCTGACCTAAC</td>
</tr>
<tr>
<td>flc6-1</td>
<td>LH1610: CTTGGCTTTTGAACAGGTAAGG</td>
<td>LH1422: TACACTATCTCTTCTCAAGAAT</td>
</tr>
<tr>
<td>emb1507-4</td>
<td>LH0168: GATCCAGGCCTGCTTTCCA</td>
<td>LH0169: GTTCAGTCCAGGCTGCTTTCCA</td>
</tr>
<tr>
<td>MSI1-TAP1 in msi1-1</td>
<td>LH1379: GTTCCAAAGCACAATTTCAGCA</td>
<td>LH1378: CAACATGGGAGGTTAATGTTAA</td>
</tr>
<tr>
<td>0.3 185 cää</td>
<td>SAIL left border T-DNA primer</td>
<td>SALK left border T-DNA primer</td>
</tr>
<tr>
<td>for dCAPS</td>
<td>TTAAG</td>
<td>LB3: TAGCATCTGAAATTCATAAACCATCTCGATAAC</td>
</tr>
<tr>
<td></td>
<td>LBb1: GCGTGGACCCTGCTGCAACT</td>
<td></td>
</tr>
</tbody>
</table>

2.1.3 Bacteria strains and plasmids

All bacteria and plasmids used in this study are listed in Table 2.3 and Table 2.4

Table 2.3 List of bacteria strains used in this study

<table>
<thead>
<tr>
<th>Bacteria strain</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli strain DH5α</td>
<td>Plasmids amplification</td>
</tr>
<tr>
<td>E. coli strain DB3.1/ccdB survivor cells</td>
<td>Gateway destination vector Amplification</td>
</tr>
<tr>
<td>BL21 Rosetta pLys</td>
<td>Protein expression</td>
</tr>
</tbody>
</table>
Table 2. 4 List of plasmids used in this study

<table>
<thead>
<tr>
<th>Name of plasmid</th>
<th>Selection in bacteria</th>
<th>Application</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pQE30-GCN5</td>
<td>Amp</td>
<td>recombinant protein expression</td>
<td>(Benhamed et al, 2006)</td>
</tr>
<tr>
<td>pGEX4T1</td>
<td>Amp</td>
<td>recombinant protein expression</td>
<td>(Amersham)</td>
</tr>
<tr>
<td>pET100-H3 (1-50)</td>
<td>Amp</td>
<td>recombinant protein expression</td>
<td>(this work)</td>
</tr>
</tbody>
</table>

2.1.4 Enzymes

Enzymes used in this study are listed in Table 2. 5

Table 2. 5 Enzymes used for DNA purification, proteomics and for immunostaining

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>proteinase K</td>
<td>In Vitro Sweden AB, Sweden</td>
</tr>
<tr>
<td>ArgC</td>
<td>Roche, Switzerland</td>
</tr>
<tr>
<td>Sequencing grade modified Trypsin</td>
<td>Promega, Switzerland</td>
</tr>
<tr>
<td>Cellulase (Onozuka) R-10</td>
<td>GTF (Fisher Scientific, Thermo Labsystems AB)</td>
</tr>
<tr>
<td>Pectinase and Pectolyase</td>
<td>Sigma-Aldrich, Germany</td>
</tr>
</tbody>
</table>

2.1.5 Antibodies

Antibodies used for Western blotting (Razi et al, 2008), Immuno-Fluorescence (IF), or for Chromatin Immuno-precipitation assay (ChIP) are listed in Table 2. 6

Table 2. 6 List of antibodies used in this study

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Origin</th>
<th>Provider</th>
<th>Cat.No</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-GST</td>
<td>Rabbit</td>
<td>Bethyl Laboratories</td>
<td>A190-122A</td>
<td>WB</td>
</tr>
<tr>
<td>Anti-His6 (2)</td>
<td>Mouse</td>
<td>Roche</td>
<td>4905318001</td>
<td>WB</td>
</tr>
<tr>
<td>Anti-Mouse IgG, HRP-linked</td>
<td>Sheep</td>
<td>Amersham Life Science</td>
<td>NXA931-1ml</td>
<td>WB</td>
</tr>
<tr>
<td>Anti-Rabbit IgG, HRP</td>
<td>Goat</td>
<td>BIO-RAD</td>
<td>170-6515</td>
<td>WB</td>
</tr>
<tr>
<td>IgG</td>
<td>Rabbit</td>
<td>Sigma-Aldrich</td>
<td>15006-10mg</td>
<td>ChIP</td>
</tr>
<tr>
<td>Anti-Histone 3, CT, pan</td>
<td>Rabbit</td>
<td>Upstate/Millipore</td>
<td>07-690</td>
<td>WB +ChIP</td>
</tr>
<tr>
<td>Anti-H3K27me3</td>
<td>Rabbit</td>
<td>Upstate/Millipore</td>
<td>07-449</td>
<td>WB +ChIP</td>
</tr>
<tr>
<td>Anti-H3K36me3</td>
<td>Rabbit</td>
<td>Diagenode</td>
<td>pAb-058-050</td>
<td>WB +ChIP</td>
</tr>
<tr>
<td>Anti-H3K4me3</td>
<td>Rabbit</td>
<td>Upstate</td>
<td>07-473</td>
<td>WB +ChIP</td>
</tr>
<tr>
<td>Anti-H3K36ac</td>
<td>Rabbit</td>
<td>ActiveMotif</td>
<td>39379</td>
<td>WB +ChIP+ IF</td>
</tr>
<tr>
<td>Anti-H3K23me1</td>
<td>Rabbit</td>
<td>ActiveMotif</td>
<td>39387</td>
<td>WB +ChIP+ IF</td>
</tr>
<tr>
<td>Anti-H3K9me2</td>
<td>Rabbit</td>
<td>Diagenode</td>
<td>pAb-060-050</td>
<td>WB +ChIP+ IF</td>
</tr>
<tr>
<td>Anti-H3K1221ac</td>
<td>Rabbit</td>
<td>Abcam</td>
<td>ab33309</td>
<td>WB+IF</td>
</tr>
<tr>
<td>Anti-Rabbit IgG - Rhodamine</td>
<td>Donkey</td>
<td>ThermoScientific/Pierce</td>
<td>31685</td>
<td>IF</td>
</tr>
</tbody>
</table>
2.2 Methods

2.2.1 Growth conditions

2.2.1.1 Bacteria

Bacteria were grown either in Luria-Bertani (LB) medium or on LB plates solidified with 0.8% agar amended with an appropriate antibiotic. *E. coli* strains were grown at 37°C. All antibiotics were prepared as 1000-fold concentrated solutions and kept at -20°C.

2.2.1.2 Plants

Seeds of *Arabidopsis thaliana* were surface sterilized by treatment with 70% ethanol for 2 min followed by 5 min with 5% sodium hypochlorite solution, and then washed 3 times with sterile distilled water. Seeds were then sown on Murashige and Skoog (MS) basal salt medium (Duchefa, Haarlem, The Netherlands). After stratification for 24 hours at 4°C, the seeds were transferred into a growth chamber containing mixed cold fluorescent and incandescent light (100-200 μmoles m⁻² s⁻¹), and grown for 10 days at 21°C, under a 16/8 h or 8/16 light-dark cycle corresponding to long day- or short day- photoperiod conditions, respectively. Plants were either planted in the soil, or directly collected for RNA extraction or ChIP assay. For nuclei and crude histone extraction, wild type *Brassica oleracea* (cauliflower) heads were obtained from a local market. Tobacco (*Nicotiana tabacum*), wheat (*Triticum aestivum*) and rice (*Oryza sativa*) were grown in greenhouse, at 22°C under long day (LD, 16 h light) photoperiods, for rice special watering conditions were applied; spruce (*Picea abies*) needles were collected from a ~15 years old wild-grown tree in a forest near Uppsala (Sweden) located at 59.81° N, 17.66° E. The flowering time of *Arabidopsis* plants was measured as the number of total rosette leaves longer than 0.5 cm at bolting for at least 14 plants.

For Zebularine (DNA methylation inhibitor) treatment of *Arabidopsis*, sterilized seeds as described above were germinated on MS medium containing 80 μM Zebularine (AH Diagnostics, Solna, Sweden). After 10 days, seedlings were collected for RNA extraction and ChIP assays. Columbia (Col-0) was used as the wild-type accession in all experiments.

2.2.2 Molecular methods

Standard molecular biology procedures (e.g. restriction enzyme digestions, plasmid DNA preparations, PCR product purification and gel electrophoresis) were performed as described in (Sambrook, et al. 1989) or according to the enzymes and kit supplier’s recommendations.

2.2.2.1 Methods for DNA

2.2.2.1.1 Plant DNA extraction with Edwards buffer

A frozen small leaf or entire seedling in a 1.5 ml reaction tube containing glass beads was ground to fine powder with a grinding machine (Silamat S5). Five hundred μl of Edwards Buffer [0.2 M Tris pH 8.0, 0.25 M NaCl, 0.025 M EDTA (ethylene diamine-tetraacetate), 0.5% SDS ] (Edwards et al, 1991) was added and mixed thoroughly to obtain a homogeneous solution. After 5 min of centrifugation at room temperature at 17,000 × g, the supernatant was collected into a clean 1.5 ml reaction tube.
DNA was precipitated by addition of 400 µl of isopropanol. After an additional centrifugation for 5 min, the supernatant was discarded and the pellet was washed with 1 ml of 70% Ethanol. The pellet was air dried, then re-suspended in 100 µl of sterile distilled water and stored at 4°C until the analysis. The genomic DNA was used for genotyping using gene specific primers and an internal T-DNA specific primer as described in Table 2.2.

2.2.2.1.2 Molecular cloning

For the cloning of the amino-terminal amino acid residues of the Arabidopsis histone 3, the desired region was amplified using cDNA produced from Col RNA (Sambrook, et al. 1989); using PhusionTM High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland). The PCR product was purified from an agarose gel using a gel extraction kit (Fermentas, Helsingborg, Sweden) according to manufacturer’s recommendations, and subcloned into a pET100/D topo plasmid according to manufacturer’s recommendations (Invitrogen, Stockholm, Sweden). The construct was confirmed by sequencing. The plasmid was subsequently transformed into BL21 for protein expression.

2.2.2.1.3 Identification of EMS-induced mutations in Arabidopsis by Illumina Deep-Sequencing

In order to identify EMS (Ethyl Methane Sulfonate) induced mutations responsible for the early flowering phenotype in Arabidopsis mutant cäö, the following procedure was employed. Mutant plants in the Col background were crossed to a diverged accession (Landsberg erecta, Ler) to generate F1 hybrid plants (Exner, 2008). F2 plants were grown under long day photoperiod conditions. Leaf samples of equal size were collected from approximately 150 plants exhibiting the mutant phenotype, pooled and subjected to bulk genomic DNA extraction using a Nucleon Phytopure genomic DNA extraction Kit (Amersham Bioscience, Uppsala, Sweden) according to manufacturer’s recommendations. Library preparation using a NEB Genomic DNA Sample Preparation Kit was performed at the functional genomic facilities at D-BSSE, Basel, Switzerland. The prepared DNA was loaded onto an Illumina Genome sequencer Ellac GA IIX, and run for 36 cycles. The obtained short reads were mapped against the TAIR9 release of the Arabidopsis genome, using the bowtie2 (Langmead and Salzberg, 2012) short read mapping software with default settings Genome-wide SNP positions and pileup information were then collected and analyzed using Next-generation EMS mutation mapping (NGM) online software (http://bar.utoronto.ca/ngm/cgi-bin/emap.cgi) (Austin et al, 2011).

2.2.2.1.4 Validation of the single nucleotide polymorphism in cäö

dCAPS markers were used for the detection of the single nucleotide polymorphism in the cäö mutant. A reverse and a forward primer with two nucleotide mismatches were designed with dCAPS Finder 2.0 online software. A fragment of 140 bp from genomic DNA of the wild-type Col and mutant cäö was amplified by classic PCR reaction with the following cycling conditions (98°C 5 s, 55°C 30 s, 72°C 40 s); the amplicon was digested without prior purification for two hours with HpaI (Fermentas) at 37°C, then loaded on a 2.5% agarose gel and separated electrophoretically. The mutated allele produces a band at 140 bp due to the alteration of HpaI recognition site in cäö, while the intact allele in Col will produce a band at 120 bp due to the digestion of the PCR product. The SNP in cäö was further validated by Sanger sequencing using LH1609 - LH1324 primers (Table 2.2).
2.2.2.2 Methods for RNA

2.2.2.2.1 RNA Extraction and RT-PCR

RNA extraction and reverse transcription were performed as described previously (Alexandre et al, 2009) with minor modification. RNA was extracted from seedlings or rosette leaves using TRIzol Reagent according to manufactures recommendations (Invitrogen). For RT-PCR analysis, 1 µg total RNA was treated with DNase I (Invitrogen) to eliminate any accidental DNA contaminations. The DNA-free RNA was reverse-transcribed using a RevertAid First Strand cDNA Synthesis Kit (Fermentas) according to manufacturer’s recommendations. Aliquots of the cDNA were used as template for PCR with gene-specific primers. For qPCR analysis, gene-specific primers and SYBR green master mix (Fermentas) were used on an IQ5 multicolor Real time PCR cycler (BIO-RAD, PA, USA). PP2A was used as reference gene. Quantitative RT-PCR was performed using three technical replicates and results were analyzed as described in (Simon, 2003). For primer sequences see (Table 2.4 and Supplemental tables).

2.2.3 Methods for proteins

2.2.3.1. Protein expression and purification from bacteria

The pET100-H3(1-50),pQE30-GCN5 and pGEX4T1 constructs were transformed into BL21pLys Rosetta cells. The transformed cells were grown overnight in 5 ml of liquid LB medium supplemented with ampicillin at (100 µg/ml). 1 ml of the bacterial cell suspension from each culture was used as an inoculum for 100 ml of LB medium supplemented with ampicillin at (100 µg/ml). The bacteria cultures were grown to an optical density, OD600 of 0.6, then 1 mM IPTG was added, and the cultures were vigorously shaken for an additional 5 hours at room temperature. The cultures were harvested by centrifugation at 5000 × g for 20 min at 4°C. The supernatants were decanted, and the pellets were frozen and stored at -80°C, until further analysis.

2.2.3.2. Purification of 6x His-tagged proteins with TALON metal affinity resin

For the purification of His-tagged expressed proteins, TALON metal affinity resin (Clontech, Stockholm, Sweden) was used. The bacteria pellet was mixed gently with cold lysis buffer, containing 1× Bug Buster (VWR International, Stockholm, Sweden), Benzonase 1 µl (25 units/ml of extraction buffer (VWR International) and complete EDTA-free protease inhibitor (Roche) and left at 4°C for 15 min until clear lysate was obtained. The total lysate was cleared by centrifugation in a cooled centrifuge for 15 min at 17.000 × g. The soluble fraction was incubated with 250 µl of 50% slurry of TALON Resin (Clontech, Stockholm, Sweden), pre-equilibrated with 1 ml of chilled lysis buffer. Beads and lysate were allowed to rotate at 4°C for 90 min. The beads were then left to settle at the bottom of a chromatography tube and the lysate to drain. Bead beds were washed with chilled washing buffer (50 mM NaH2PO4, 300 mM NaCl, 20 mM imidazole), then 0.5 ml of chilled elution buffer (50 mM NaH2PO4, 300 mM NaCl, 250 mM imidazole) was applied to elute tagged proteins. Aliquots from the eluates were taken to measure protein concentration with Bradford assays (Bradford and Williams, 1976), and purity and protein size were checked by SDS-PAGE and Western blotting, respectively. The remaining of the purified protein was stored at -20°C until further analysis.
2.2.3.3. Purification of GST-tagged proteins with affinity chromatography

For the purification of GST-GCN5 or GST alone, Glutathione Sepharose 4B resin was employed. The bacteria pellet was mixed gently with cold lysis buffer (as above) and left at 4°C for 15 min until a clear lysate was obtained. The total lysate was cleared by centrifugation in a cooled centrifuge for 15 min at 17,000 × g. The soluble fraction was incubated with 250 µl of 50% Glutathione Sepharose 4B (Amersham Life Science, Stockholm, Sweden), pre-equilibrated with 1 ml of chilled binding buffer [20 mM Tris pH8.0, 120 mM NaCl, 0.1% Triton X-100]. Beads and lysate were allowed to rotate at 4°C for 90 min. The beads were then left to settle at the bottom of a chromatography tube and the lysate to drain. After beads washing, 0.5 ml of chilled Glutathione elution buffer (10 mM GSH, 50 mM Tris-HCl, pH8.0) was applied to elute tagged proteins. Aliquots from the eluates were taken to measure protein concentration with Bradford assays (Bradford and Williams, 1976), and purity and protein size were checked by SDS-PAGE and Western blotting, respectively. The remaining of the purified protein was stored at -20°C until further analysis.

2.2.3.4 Histone extraction from plants

Approximately 3 g of inflorescences from Brassica oleracea (cauliflower) or Arabidopsis, or adult leaves from Arabidopsis, wheat, rice or spruce were ground to a fine powder and homogenized for 15 min in histone extraction buffer (0.25 M sucrose, 1 mM CaCl₂, 15 mM NaCl, 60 mM KCl, 5 mM MgCl₂, 15 mM PIPES, pH 7.0, 0.5% Triton X-100 including protease inhibitors cocktail (Roche) and 10 mM sodium butyrate). After centrifugation at 4°C for 20 min and 4500 × g, pellets were dissolved in 0.1 M H₂SO₄, and left to shake for overnight at 4°C. After centrifugation for 10 min at 17000 × g, total histones were precipitated from the supernatant with concentrated Trichloroacetic acid to a final concentration of 33%. The histone pellet was washed twice with an acetone-0.1% HCl mixture and once with acetone, air dried and then dissolved either in 1× Laemmli buffer and subjected to SDS-PAGE followed by Western blotting, or subsequently fractionated by RP-HPLC.

2.2.3.5 Histone fractionation by RP-HPLC and post translational modification identification by Tandem Mass Spectrometry

To identify the different post translational modifications occurring on the N-terminal tail of the histone H3, bulk histones were prepared as described above. The pellet was then dissolved in 100 µl of H₂O, sonicated for 1 min, and centrifuged for 1 min at 10,000 × g. The supernatant was loaded on an ECLIPS XDB-C8 column (4.6x150 mm) (Agilent, Switzerland) connected to an Agilent HP1100 binary HPLC system. Histones were separated and eluted with the following gradient: 0-5 min, 0% solvent B; 5-15 min, 0-35% Buffer B; 15-25 min, 35% Buffer B, 25-75 min, 35-65% Buffer A. Buffer A was 5% acetonitrile (ACN) in 0.1% TFA and Buffer B was 90% ACN in 0.1% TFA (Tweedie-Cullen et al, 2009). Aliquots of the fractionated histones were separated on 15% SDS-PAGE and blotted with a rabbit anti-H3 antibody. The fractions containing H3 were combined, dried under vacuum and dissolved in 50 mM ammonium bicarbonate pH 8.0 to be used for mass spectrometry analysis. The histones were first reduced with 10 mM dithiothreitol (DTT) for 30-60 min, alkylated with 40 mM iodoacetamide for 1h and then digested for 2 h at 37°C with Chymotrypsin (enzyme:substrate ratio 1:50) (Promega, Switzerland), semi-trypsin (enzyme:substrate ratio 1:200) (Promega, Switzerland) or with Arg C (1:50) (Roche, Basel, Switzerland). TFA was added to a final concentration of 0.1% to stop the digest.
Samples were concentrated and desalted with C18 ZipTips (Millipore, Zug Switzerland) according to the following procedure; the Zip-Tip was washed twice with 10 μl of 50% acetonitrile and 0.1% trifluoroacetic acid. Thereafter, the tip was equilibrated twice with 10 μl of 0.1% trifluoroacetic acid. Peptide binding was performed by pipetting the solution 3-10 times through the filter matrix. Afterwards, the tip was washed four times with 10 μl of 0.1% trifluoroacetic acid; finally, the peptides were eluted with 10 μl of 50% acetonitrile and 0.1% trifluoroacetic acid, and subsequently stored at -20°C prior to mass spectrometry analysis. Samples were analyzed on an LTQ-Orbitrap mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) interfaced with a nano electrospray ion source, at the Proteomics unit in Functional Genomic Center (FGCZ) Zurich, Switzerland.

2.2.3.6 Mass Spectrometry database analysis

Database searches of MS/MS spectra using Mascot version 2.1.10 were performed using a TAIR9 Arabidopsis protein database. Modifications used for searches included carbamidomethyl (cystein (C), fixed, non-histone search), oxidation (methionine (M), variable histone and non-histone modification search), acetyl (N-term protein and Lysin (K), variable, histone searches), mono-, di-, and trimethylation (arginine (R) and lysin (K), variable, histone searches). Only strictly tryptic peptides with a maximum of 1 (fully tryptic digests) or 3 (semitryptic histone digests) missed cleavage sites were allowed in database searches. For Orbitrap measurements, the monoisotopic masses of +1, +2 and +3 charged peptides were searched with a peptide tolerance of 6 ppm and an MS/MS tolerance of 0.5 Da for fragment ions. Positive identification of acetylated and methylated peptides was performed using a variety of strict criteria including manual, individual inspection of spectra. Only **bold-red**, rank 1 peptides with a Mascot expect values of less than 0.05 were considered. In addition, spectra from all peptides derived from histones were manually validated for site placement. The confirmation of modification sites was primarily based on the presence of site-specific singly or doubly charged b and y type fragment ions (b and y ions generated by cleavages between two potentially modified residues). Relative intensities of essential diagnostic fragment ions were checked in MS/MS spectra.

2.2.3.7 Protein separation on SDS-PAGE and Western-blotting

The protein samples were loaded on 12% SDS-polyacrylamide gels for higher molecular weight proteins or on 15% gels for histones. The proteins were blotted onto pre-wetted PVDF membrane. The transfer was left to occur at 15 V for 45 min on a Trans blot SD (BIO-RAD, Germany). The membrane was subsequently blocked with 5% nonfat milk in TBST at room temperature for 1h.

After 3 times washing with TBST, for 5 min each, the membrane was incubated with the desired primary antibody at 4°C for 3 h to overnight. The membrane was washed for additional 3 times with TBST and incubated with horseradish peroxidase labeled secondary antibodies for an additional 1h. For protein detection, the ECL Western blotting kit was used (Amersham Life Science, Stockholm, Sweden) and documented on an Fuji LAS3000 camera visualization system.

2.2.3.8 Dot blotting

To test the presence of H3K36ac and H3K23me1 in cauliflower and in Arabidopsis inflorescences or seedlings, 1μg of acid extracted histones dissolved in water was blotted onto a Nitrocellulose membrane Hybond ECL™ (Amersham Life Science, Stockholm, Sweden) and left to air dry for approximately 1h. The membrane was then blocked with 5% nonfat milk in TBST and processed as described for Western blots.
2.2.4 Methods for chromatin

2.2.4.1 Crosslinked Chromatin Immunoprecipitation assays

ChIP assays were performed according to (Exner, et al. 2009). Briefly, 1.5 g of ten-day-old entire seedlings were harvested and crosslinked in formaldehyde for 10 min, the crosslinking was stopped by the addition of 125 mM Glycin. Seedlings were washed with distilled water, and subsequently frozen in liquid nitrogen and ground to fine powder, then dissolved in nuclei extraction buffer. The purified nuclei were sonicated in 500 µl lysis buffer using a Bioruptor NextGen (Diagenode, Belgium) for (20s, on; 45s, off; 14 cycles, in shilled water bath). The cleared chromatin was diluted in nuclear dilution buffer and pre-cleared at 4°C for 1h with Protein A agarose beads pre-absorbed with sheared salmon sperm DNA (Millipore). One tenth of the precleared chromatin preparation was kept as Input control, and the chromatin was precipitated at 4°C for overnight with 1 µg of the desired antibodies (Table 2.6).

The immune serum was collected by the addition of Protein A agarose beads. After several washings, the eluted chromatin was reverse cross linked, the DNA was precipitated and analyzed by quantitative real-time PCR (qPCR) using SYBR green master mix (Fermentas) and gene-specific primers on a IQ5 multicolor Real time PCR thermo cycler (BIO-RAD). Quantitative PCR was performed with three technical replicates and presented as percent of input as described in (Simon, 2003).

2.2.4.2 Native ChIP-seq

Native Chromatin Immunoprecipitation (ChIP) was performed as described in (Bernatavichute et al, 2008) by Shu Huan (ETH Zurich). Crude nuclei extracts, from leaf number six of plants grown for 35 days at SD conditions harvested at zeitgeber time (ZT) =7 (i.e. 7 hours after light on), were produced by treating 100 mg of ground frozen plant material in Nuclei Extraction Buffer [NEB; 20 mM PIPES-KOH pH 7.6, 1 M hexylene glycol, 10 mM MgCl2, 0.1 mM EGTA, 15 mM NaCl, 60 mM KCl, 0.5% Triton-X, 5 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche)] for 15 min at 4°C. The homogenate was filtered through Miracloth (Calbiochem,Germany) and a pellet was harvested by centrifugation for 10 min at 1500×g at 4°C. Isolated nuclei were washed one time in MNase buffer [50 mM Tris-HCl pH8.0, 10 mM NaCl, 5 mM CaCl2, and EDTA-free protease inhibitor cocktail (Roche)] and were treated with 1.3 µl of RNase A, 30 µg/µl (Sigma-Aldrich, USA) and used for Micrococcal Nuclease (New England BioLabs, NEB, USA) digestion for 4 minutes (final concentration 0.2 U/µl) in MNase buffer. The reaction was terminated by addition of 10 mM EDTA. The supernatant after centrifugation was collected as phase 1 chromatin preparation, and the supernatant from a second centrifugation after thirty minutes treatment of the pellet with S2 buffer [1 mM Tris- HCl pH 8, 0.2 mM EDTA, and EDTA-free protease inhibitor cocktail (Roche)] was collected as phase 2 chromatin preparation. The two phases of chromatin preparations were combined and the NaCl concentration was adjusted to 50 mM. The majority of the chromatin was mononucleosome size (data not shown). Histone H1 was depleted by incubating the chromatin preparation with Sephadex C25-CM resin (Pharmacia, Sweden) for 1h at 4°C (Thorne et al., 2004). The Triton-X concentration in the chromatin preparation was brought up to 0.1%, and chromatin preparation was pre-cleared using non-immune rabbit IgG (see below) and Dynabeads Protein A (Invitrogen).

One tenth of the precleared chromatin preparation was kept as Input control, and ¼ was used for each immunoprecipitation with 2.5 µg antibody [anti-H3 antibody #07-690 (Upstate/Millipore); anti-H3K36ac #39379 (ActiveMotif, USA); non-immune rabbit IgG #15006 (Sigma-Aldrich, Germany) and collected with Dynabeads Protein A (Invitrogen).
After washing, beads were re-suspended in TE buffer [10 mM Tris-HCl, pH 7.5, 1 mM EDTA], and DNA was extracted using conventional phenol-chlorophorm extraction and an ethanol/salt precipitation protocol (Sambrook, et al. 1989). ChIP was performed in three technical replicates. The sequencing was performed at the Next Generation Sequencing Unit in the Functional Genomic Center (FGCZ) Zurich, Switzerland.

2.2.5 Histone Acetyl Transferase activity Assay

Liquid HAT assays were performed in a final volume of 30 µl. One µg of purified His-tagged H3 tail (1-50) was mixed with the purified recombinant GST-GCN5 protein in HAT buffer [0.25 µg/µl Acetyl-CoA (Sigma-Aldrich, Germany), 50 mM Tris-HCl (pH 8.0), 10% glycerol (v/v), 1 mM dithiothreitol, complete EDTA-free protease inhibitor (Roche), and 10 mM sodium butyrate]. Acetylation reactions were incubated for 60 min at 30°C with gentle shaking and were stopped by addition of 2×SDS sample buffer and heating at 100°C for 5 min. The reaction product was resolved on 15% SDS-polyacrylamide gels, followed by Western blotting. The acetylated H3 was identified by using anti-H3K36ac antibodies (#39379, ActiveMotif, USA), and the loading control was visualized with anti-H3 antibodies (#07-690, Upstate/Millipore).

2.2.6 Immunostaining

Immunostaining was performed according to (Jasencakova et al, 2000). Briefly, the root tips of four-day-old Arabidopsis seedlings were excised under microscope and fixed for 20 min in ice-cold 4% (w/v) paraformaldehyde in MTSB buffer [50 mM PIPES, 5 mM MgSO₄, and 5 mM EGTA, pH 7.9]. Seedlings’ root tips were subsequently washed three times for 5 min in MTSB, and then digested at 37°C for 10 min with a PCP enzyme mixture [2.5% pectinase, 2.5% cellulase Onozuka R-10, and 2.5% Pectolyase Y-23 (w/v) dissolved in MTSB]. After washing three times with MTSB, root tips were squashed on a glass slides with a cover slip and immediately frozen in liquid nitrogen. After gentle removal of the cover slip, the slides were covered with 100 µl of blocking solution (MTSB containing 3% BSA) for 1 hour at 4°C.

The blocking solution was discarded, and digested cells were incubated with anti H3K36ac (#39379, ActiveMotif, USA) or anti H3K23me1 (#39387, ActiveMotif, USA) primary antibodies in MTSB containing 1% BSA and 0.1% Tween 20 and kept in a humid condition for overnight. Samples were washed three times for 5 min in MTSB and further incubated for 1-2h at 4°C with Rhodamine-conjugated anti-rabbit IgG antibodies (Thermo Scientific) diluted in MTSB containing 1% BSA and 0.1% Tween 20. The slides were washed three times for 5 min with MTSB; DNA was counterstained with DAPI in Vectashield mounting medium (Vector Laboratories). Fluorescence signal detection and documentation was performed with the help of a Leica DMI 4000 microscope. Images were taken and electronically overlayed resulting in a blue autofluorescence and red histone modification-specific fluorescence. Images were merged and processed using Photoshop and ImageJ.

2.2.7 Sequence alignment and phylogenetic analysis

Amino acid sequences of H3 and Brr2 from higher organisms were obtained using PSI-BLAST searches at [http://www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/). Representative organisms from the different eukaryote kingdoms were selected, and their amino acid sequences were obtained from protein data bases at NCBI ([http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Search&db=pubmed](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Search&db=pubmed)). Amino acid sequences were aligned using ClustalW implemented in MEGA5 (Tamura et al, 2011). ATPase and helicase domains were predicted using SMART protein prediction software (Letunic, et al. 2009). A multiple alignment was performed with gap opening penalty and gap extension penalty of 10 and 0.2, respectively. The protein weight matrix used was Gonnet 250, with a delay divergent cutoff of 30%.
Evolutionary analyses were conducted in MEGA5 (Tamura, et al. 2011) software, gaps were automatically removed and a bootstrap Neighbor -joining tree was calculated with a bootstrap trial number of 1000. Visualization of the amino acids alignment and generation of a phylogenetic tree were done with GenDoc, and MEGA5, respectively. The amino acid sequences in the FASTA format are available in Supplemental tables.
The BRR2a splicing protein is involved in Arabidopsis flowering time

Reference:
Manuscript in preparation

Authorship:
I performed plant analysis, carried out most of the molecular and genetic studies and wrote the manuscript.
The BRR2a splicing protein is involved in Arabidopsis flowering time

Mahrez, Valid1, Li, Jinghai2, Exner, Vivien1, Siretsky, Alexey2, Gruissem, Wilhelm1 and Hennig, Lars1,2,3

Affiliations

1 ETH Zürich, Department of Biology, Universitätsstrasse 2, CH-8092 Zürich, Switzerland.
2 Swedish University of Agricultural Sciences, Department of Plant Biology and Forest Genetics and Linnean Center for Plant Biology, PO-Box 7080, SE-75007 Uppsala, Sweden.
3 Science for Life Laboratory, SE-75105 Uppsala, Sweden

SUMMARY

The switch from the vegetative to the reproductive phase is an important developmental step for flowering plants. Several pathways control the flowering time in Arabidopsis through transcriptional and posttranscriptional gene regulation. mRNA processing has emerged as a critical step for flowering time control in plants. However, little is known about the molecular mechanisms linking RNA splicing to flowering time. In a mutant screen for Arabidopsis thaliana mutants with an early flowering phenotype we identified an allele of BRR2a. BRR2 proteins are components of the spliceosome and highly conserved throughout the eukaryotes. Arabidopsis BRR2a is ubiquitously expressed in all tested tissues. We found that BRR2a is involved in transcript processing of flowering time genes, most notably FLC. Mutation on threonine 895 of BRR2a caused defects in FLC splicing and greatly reduced FLC transcript levels. Reduced FLC expression allows increased transcription of FT and SOC1 leading to early flowering in both short and long days. We propose that BRR2a is in particular needed for efficient splicing of genes with long and structured introns such as in FLC and that this function of BRR2 proteins is conserved between yeast, plants and animals.
INTRODUCTION

The switch from the vegetative to the reproductive phase is an important developmental step for flowering plants. The timing of this step is regulated by numerous factors, including endogenous cues and environmental stimuli. Three main genetic pathways, the photoperiod, vernalization and autonomous pathway have been described to control flowering time in the model plant Arabidopsis. The different flowering pathways converge at the activation of a common set of genes that promote flowering and are known as floral integrators, namely SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1), FLOWERING LOCUS T (FT) and LEAFY (LFY), and at the repression of the major flowering repressor FLOWERING LOCUS C (FLC) (Srikanth and Schmid, 2011).

The daily light duration period is sensed by the photoperiod pathway. In temperate climates, most plants flower earlier in long day (LD) than in short day (SD) conditions (Stinchcombe et al., 2004). The photoperiod pathway functions to activate FT expression in leaves. FT protein is a major mobile flowering inducing signal and moves through the phloem into the shoot apical meristem (SAM) where it induces a change in identity and the switch to flowering (Andres and Coupland, 2012). In many accessions of Arabidopsis, prolonged cold exposure strongly shortens the time to flowering. The sensing pathway is known as vernalization pathway. Vernalization-dependent accessions of Arabidopsis carry active alleles of FLC and FRI, whereas vernalization-independent accessions contain inactive alleles of either one or both genes. Vernalization acts epigenetically to promote flowering by repressing the expression of FLC via the establishment of stable repressive histone marks on its promoter and gene body (Zografi and Sung, 2012). The autonomous pathway is known to promote flowering independently from environmental signals (Quesada et al., 2005). Mutants in the autonomous pathway are extremely late flowering, but vernalization suppresses their phenotype. The autonomous pathway genes act as FLC repressors, and they belong to two main subfamilies: (i) chromatin modifiers (He and Amasino, 2005) such as FLOWERING LOCUS D, FLD (He et al., 2003), FVE (Ausin et al., 2004), RELATIVE OF EARLY FLOWERING 6 (Noh et al., 2004) and HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY (Han et al., 2007) and (ii) putative RNA binding proteins (RBP) such as FCA (Simpson et al., 2003), FPA (Schomburg et al., 2001), FY (Simpson et al., 2003), FLOWERING LOCUS K (FLK) (Lim et al., 2004) and LUMINODEPENDANCE (LD) (Lee et al., 1994). Members of the last subfamily carry specific motifs facilitating recognition and binding of RNA molecules.

Although it was shown that some of these RBPs bind to FLC transcripts and reduce their abundance, the mechanism(s) behind this effect remain unclear (Lorkovic, 2009). It should be noted that genetic studies on the autonomous pathway have established that this pathway is a group of partially independently acting genes rather than a classical linear genetic pathway (Quesada et al., 2005).

Similar to many other plant genes, transcripts of most flowering-related gene are extensively processed, most notably spliced. Splicing is an accurate, dynamic and highly regulated process that removes non-coding introns from pre-mRNAs to form mature mRNA, which can be subsequently translated into proteins. The spliceosome is a macromolecular complex, consisting of five highly conserved small nuclear ribonucleoprotein particles (snRNPs; U1, U2, U4, U5 and U6) and a large number of proteins that stabilize the spliceosome complex. The removal of the introns occurs through two trans-esterification reactions during the activation step of the splicing (Moore and Sharp, 1993). The splicing reaction can be functionally divided into several steps, including spliceosome assembly, activation, catalysis, and disassembly of the spliceosomal machinery. Based on in vitro studies using splicing systems of mammalian cells and budding yeast, models about the molecular mechanism of splicing and the dynamics of the spliceosomal protein interactions and rearrangements were described (Wahl et al., 2009). During the activation step, two DExD/H-box RNA helicases, Prp28 (Staley and Guthrie, 1999) and Brr2p (Raghunathan and Guthrie, 1998; Kuhn et al., 2002) are known to play key roles. The DEXD/H-box RNA helicases belong to a large, highly conserved protein family.
These proteins with their conserved helicase domain play roles in all biological processes related to RNA metabolism, using energy from ATP hydrolysis. DEXD/H-box RNA helicases are well characterized and studied for their role in the splicing process (Silverman et al, 2003). Thanks to the known structures of some of these helicases in the spliceosome context in animals and yeast, the molecular organization during the catalytic activation of the spliceosome is well understood (Pena et al, 2007; Zhang et al, 2009; Hacker et al, 2008a). Furthermore, using mutagenesis, the functional importance of the amino-terminal helicase domain on survival of both yeast and mammalian cells was revealed (Noble and Guthrie, 1996; Raghunathan and Guthrie, 1998).

RNA processing is much less well studied in plants than in animals and yeast. However, during the last decade, the functional role of transcript processing in plants had received some attention. Several lines of evidence support a connection between RNA processing and flowering time control both in Brassicaceae (Eckardt, 2002; Herr et al, 2006, Wu et al, 2012, Sonmez et al, 2011) as well as in other plant species (Bassett et al, 2000; Eckardt, 2002; Du et al, 2006).

The identified proteins were involved in either the pre-mRNA processing through the 3’ end polyadenylation or 5’ capping. However, little is known about the possible effect of key proteins of the spliceosomal complex on the control of flowering.

Here we describe an early flowering allele of Arabidopsis BRR2a, a major protein in the splicing process. This protein is highly conserved among eukaryotes; its mutation at the threonine on position 895 is associated with an early flowering phenotype as well as severe alterations in FLC transcript processing and mRNA levels.

**MATERIALS AND METHODS**

**Plant material and growth conditions**

The Arabidopsis thaliana wild-type and T-DNA insertion lines were in the Columbia (Col) background. FRI in Col (Johanson et al, 2000), msi1-tap1 (Bouveret et al, 2006) and flc-6 (Bouveret et al, 2006) were described before; emb1507-4 (NASC ID: N16092) was obtained from the Nottingham Arabidopsis Seed Stock Centre. The EMS-mutated allele căö in the msi1-tap1 background was isolated in a mutant screen that was described before (Exner et al, 2009). For further characterization, căö in the msi1-tap1 background was backcrossed into Col.

Seeds from all plants were sown on 0.5× basal salts Murashige and Skoog (MS) medium (Duchefa, Haarlem, The Netherlands), stratified at 4°C for 1 day, and allowed to germinate in growth chambers at 20°C for 10 days under LD (16 h light) or SD (8 h light) photoperiods. Next, healthy plantlets were planted in soil and grown in growth chambers under the same conditions. Scoring of the number of total rosette leaves longer than 0.5 cm at bolting was performed for at least 14 plants in order to assess the flowering time at given conditions.

**Mapping by Illumina Deep-Sequencing**

A mapping population was established by crossing căö with the polymorphic ecotype Ler, and total genomic DNA was extracted from 150 F2 plants presenting the mutant phenotype using the Nucleon Phytopure genomic DNA extraction Kit (Amersham Bioscience, Uppsala, Sweden). After library preparation using standard Illumina protocols, the DNA was loaded onto an Illumina Genome sequencer GA IIx and run for 36 cycles at D-BSSE (ETHZ, Basel, Switzerland). The obtained short reads were mapped against the TAIR10 release of the Arabidopsis genome using Bowtie 2 (Langmead et al, 2009).
Genome-wide SNP positions and pileup information were then collected and filtered as recommended in the Next-generation EMS mutation mapping (Langmead et al, 2009) software (bar.utoronto.ca/ngm/cgi-bin/emap.cgi) (Austin et al, 2011). dCAPS markers were used for the detection of the single nucleotide polymorphism in the cãö mutant. Primers, which are listed in Table 2.2, were designed using dCAPS Finder 2.0 (Neff et al, 2002). The amplified fragments from genomic DNA of the wild type Col and mutant cãö were digested with HpaI (Fermentas, Helsingborg, Sweden) and loaded on a 2.5% agarose gel. The SNP in cãö was further validated by standard Sanger sequencing using primers LH1609: CTTGAAGGAAGATAGTGTAACTCGT and LH1324: CCGAATGTATCAGGTCAGCTCTT.

RNA isolation and RT-qPCR

RNA extraction and reverse transcription were performed as described previously (Alexandre et al, 2009) with minor modifications. The DNA-free RNA was reverse-transcribed using a RevertAid First Strand cDNA Synthesis Kit (Fermentas, Helsingborg, Sweden) according to manufacturer’s recommendations. Aliquots of the generated cDNA were used as template for PCR with gene-specific primers (Supplemental Tab. S1, S2). Quantitative PCR was performed using gene-specific primers (Supplemental Tab. S3) and SYBR green (Fermentas, Helsingborg, Sweden) on an IQ5 multicolor Real time PCR thermo cycler (BIO-RAD, PA, USA). qPCR reactions were performed in triplicate; gene expression levels were normalized to a PP2A control gene, and results were analyzed as described (Simon, 2003).

Sequence alignment and phylogenetic analysis

Protein sequences of BRR2a homologues were obtained using PSI-BLAST searches, representative organisms from the different eukaryote kingdoms were selected, and their BRR2 amino acid sequences retrieved from protein databases at NCBI. ATPase and helicase domains were predicted using SMART protein prediction software (Letunic et al, 2009). Amino acid sequences were aligned using ClustalW implemented in MEGA5 (Tamura et al, 2011). Evolutionary analyses were conducted in MEGA5 software, and a bootstrap Neighbor-Joining Tree was calculated for 1000 bootstrap trials. The list of proteins used for the alignment is given in Supplemental Tab. S5.

Analysis of intron secondary structure

Prediction of intron secondary structure was essentially performed as described (Meyer et al, 2011). A total of 175'508 introns were extracted from the Arabidopsis TAIR10 genome. To predict branch points (BP), introns were scanned for the YURAY motif (Marquez et al, 2012), up to 200 nt upstream from the 3’ splice site (3’ SS). Such a long distance was chosen also in (Meyer et al, 2011). It is plausible to use this long distance, because many Arabidopsis introns do not have a consensus BP sequence within the common 30 bp close to the 3’SS and because recent data suggest that at least in yeast the distance between BP and 3’SS can be much larger than previously thought (Meyer et al, 2011). The YURAY motif closest to the 3’ SS was predicted as BP. For secondary structure predictions of entire introns or only BP to 3’ SS sequences, RNAfold (Hofacker, 2009) was used with default parameters.
RESULTS

The cäö mutant causes early flowering

To better understand the molecular mechanisms underlying the control of the floral transition by MS11 (Bouveret et al, 2006), a mutant screen for suppressors of the late flowering phenotype of msil-tap1 plants was performed (Exner, 2008). After mutagenesis with ethylmethanesulfonate (Grillo et al, 2013), flowering time was assessed under LD photoperiods for 1045 M2 families. Eleven families had shortened the vegetative phase of msil-tap1. For six of them the phenotype was confirmed in subsequent generations (Exner, 2008). One of these six contained a mutation in LHP1 (Exner et al, 2009); the other mutant, was described to have a defect in CRY1 (Exner et al, 2010). One of the remaining uncharacterized suppressor mutants, which was called chrottapöische (cäö) (Swiss German for dandelion) because of its increased leaf serration, was chosen for further investigation.

To test whether the cäö phenotype can be observed independently of the msil-tap1 background, cäö was backcrossed into Col, and flowering time was measured. Under LD conditions, msil-tap1 flowered much later than Col confirming earlier observations (Bouveret et al, 2006), while cäö msil-tap1 flowered similar to Col demonstrating the suppression of the msil-tap1 late flowering phenotype (Fig. 3.1 A). Finally, cäö flowered earlier than both Col and cäö msil-tap1 (Fig. 3.1 A). The flowering time data show that the effect of cäö does not require the msil-tap1 background.

In addition, the rosette leaf numbers are much more similar to an additive than an epistatic genetic interaction, strongly suggesting that the gene that is defect in cäö and msil-tap1 affect flowering through independent genetic pathways. Therefore, cäö back-crossed into Col was used in all subsequent experiments.

Figure 3.1 cäö is an early flowering mutant in Arabidopsis

(A) Flowering time of wild type (Col), msil-tap1, cäö and cäö msil-tap1 in number of total rosette leaves under LD. (B) Flowering time of wild type, cäö and msil-tap1 in number of total rosette leaves under SD.

To further characterize the flowering time phenotype of cäö, the mutant was grown together with msil-tap1 and Col under SD conditions. Under these conditions, cäö differed most strongly from Col and msil-tap1 with three and six times fewer leaves formed, respectively (Fig. 3.1 B). Taken together, the mutation in cäö had considerably reversed the late flowering phenotype of msil-tap1 but causes early flowering even in a wild-type background, most strongly under SD photoperiods.
Development of *cāō* is altered

In addition to the early flowering, the *cāō* plants presented some other developmental defects. Leaves of *cāō* were serrated and small (Fig. 3.2 A, D) contributing to smaller and more compact rosettes (Fig. 3.2 B, C). Fruits of *cāō* were smaller than in Col (Fig. 3.2 E). In order to investigate the reason for smaller siliques size, 11 mature siliques from self-pollinated Col and *cāō* plants were collected, opened under a stereo-microscope, and the number of viable seeds was scored. In wild type Col, 548 (98%) of the 560 analyzed seeds had developed normally.

In contrast, only 293 (63%) of 464 seeds from self-pollinated, homozygous *cāō* mutants had developed normally revealing an important function of the gene mutated in *cāō* during seed development (Fig. 3.2 E, F).
Figure 3.2 Developmental alterations in cää

(A) Silhouette of the sixth rosette leaf from Col (left) and cää showing the serrated margin of the cää leaf; scale bar: 2 cm. (B) Rosette morphology of Col and cää plants at time of bolting; scale bar: 5 cm. (C) Rosette diameter of Col and cää (mean ± SD; n = 16). (D) Leaf morphology of Col and cää in 20 days old plants; scale bar: 1 cm. (E) Homozygous cää mutants show altered fruit development; scale bar: 20 mm. (F) Open Col siliques show full seed set. (G) Open cää siliques show much reduced seed set.
caö encodes an ATP-dependent RNA helicase protein

To determine the inheritance mode of the mutation in caö, F2 plants from a caö × Col cross were grown at LD conditions, and the number of the plants with caö phenotype was scored. Out of 74 plants 18 (~24%) had a caö phenotype. This segregation is most consistent with a single recessive allele.

In order to isolate the causative mutation in caö, a mapping population was created by crossing caö in Col with the highly polymorphic accession Ler. F2 plants were grown under LD conditions, and plants with the caö phenotype were selected.

![Graphs showing genome-wide SNP frequencies](image)

**Figure 3.3 Genome-wide SNP frequencies plotted as a function of chromosomal position**

SNPs frequencies in sequencing reads were plotted along each chromosome using a bin size of 250 kb. SNPs are created by Ler reads. A non-recombinant region on the right arm of chromosome 1 with very few Ler reads is indicated by a brace. (Chro, abbreviation for Chromosome). Histograms are output of the Next-generation mapping tool.
Pooled genomic DNA was sequenced at the ETH Department of Biosystems Science and Engineering, Basel, Switzerland. The obtained Illumina reads were mapped to the TAIR10 release of the Arabidopsis Col-0 genome using bowtie2 (Langmead et al, 2009), and data were analyzed with the Next-generation mapping tool (Austin et al, 2011).

SNP frequencies in the mapping population, binned at 250 kb intervals, were evenly distributed along chromosomes. However, there was an obvious non-recombinant region, which can be recognized as a SNP “desert” on the left arm of chromosome 1 (Fig. 3.3).

Further statistical analysis of SNP distributions on chromosome 1 using the Next-generation mapping tool revealed a candidate peak that most likely would contain the causative mutation (Fig. 3.4). This region contained only one mutation that was represented by all covering reads. The mutation was at position 7307231 in the \( \text{AT1G20960} \) gene.

**Figure 3.4 SNP localization from the NGM web application**

Screenshot of the final stage of region selection and SNP annotation. The sharp delimited peak, at position 7307231 corresponds to the position of a mutation in \( \text{AT1G20960} \).

The SNP was localized at 2685 bp downstream of the ATG, in the second exon of the open reading frame of \( \text{AT1G20960} \). The mutation is a transition of Guanine to Adenine (G to A) on the reverse DNA strand. The mutated base pair generated a missense mutation, where the polar amino acid threonine at position 895 was altered to the apolar isoleucine (Tab. 3.2). To confirm the identified mutation, a dCAPS (Derived cleaved amplified polymorphic sequences) molecular marker was applied, in which \( HpaI \) cuts only the Col but not the predicted mutant sequence. The results confirmed the presence of the G to A mutation in the \( \text{AT1G20960} \) locus (Fig. 3.5 A). For further confirmation, amplicons from Col and \( \text{cãö} \) genomic DNA were Sanger sequenced using a gene-specific primer pair. This analysis confirmed also the transition from G to A, at 3021 bp downstream of the ATG start codon (Fig. 3.5 B).

**Table 3.2 SNP annotations on left arm of Chromosome 1**

<table>
<thead>
<tr>
<th>Chro</th>
<th>Position</th>
<th>Ref base</th>
<th>SNP base</th>
<th>Depth</th>
<th>Locus</th>
<th>Strand orientation</th>
<th>Ref. codon</th>
<th>SNP codon</th>
<th>AA change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7307231</td>
<td>G</td>
<td>A</td>
<td>13</td>
<td>( \text{AT1G20960} )</td>
<td>-</td>
<td>ACT</td>
<td>ATT</td>
<td>T&gt;I</td>
</tr>
</tbody>
</table>
Figure 3.5 Confirmation of the gene mutation by dCAPS and by Sanger Sequencing

(A) Gel electrophoretic separation of PCR products (left undigested, right after HpaI digestion). The PCR digestion products were separated by electrophoresis on a 2.5% agarose gel. (B) PCR amplified fragments from Col and cāō genomic DNA were sequenced using a gene-specific reverse primer for a region where the mutation is localized. Sequences were aligned to Col and brr2a-2, and the corresponding amino acids sequences are also listed.

The gene AT1G20960 is 8503 bp long and consists of five exons (Fig. 3.6 A). This locus was previously identified among 250 others in a large-scale screen for loss-of-function phenotypes of indispensable genes and was named EMB1507 as a homozygous mutant allele was embryonic lethal (Tzafrir et al, 2004). No further characterization was reported so far.

Sequence analysis revealed that the AT1G20960 protein is an orthologue of the yeast Brr2p (Bad response to refrigeration 2 protein), and thus was named BRR2a (Liu et al, 2009). Two other Arabidopsis paralogues are BRR2b (At2g42270) and BRR2c (At5g61140). Because BRR2 is the original name given to these proteins and cāō is the second allele described for this gene, we will from now on refer to cāō as brr2a-2 and to the mutant protein as BRR2a-T895I.
Figure 3.6 BRR2a DNA and protein structure

Schematic representation of (A) the BRR2a gene with exons (black boxes), introns (black lines), untranslated regions (grey boxes), and the position of the mutation on the DNA and protein (red star). (B) the protein domain structure of BRR2a. A detailed description of the protein domains can be found in the main text.

Arabidopsis BRR2a encodes a putative protein of 2171 amino acids with a predicted molecular mass of about 247 kDa and an estimated pI of 5.5. Homologs of this protein are known as Brr2p in S. cerevisiae and U5-200kD in humans. BRR2a belongs to the DEAD/DEXH box family of ATP-dependent RNA helicases with a unique N-terminal domain and two consecutive helicase cassettes (with a DExD/H domain), each followed by a Sec63 domain (Fig. 3.6 B) (Pena et al. 2009; Zhang et al. 2009). BRR2 proteins are integral components of the U5 small nuclear ribonucleoprotein (snRNP) and are essential for splicing through their contribution to the activation and recruitment of spliceosome complex components (Raghunathan and Guthrie, 1998).

Genetic confirmation of the mutated allele in brr2a-2 mutants

To test whether the early flowering phenotype is due to the disruption of BRR2a, an allelism test was performed. Heterozygous emb1507-4 null mutant plants were used to pollinate homozygous brr2a-2 plants. Eleven F1 seeds germinated on MS plates and were transferred to soil. Six of the 11 plants carried the emb1507-4 allele, closely matching the expected 50%. Importantly, all plants with the emb1507-4 allele but none of the plants without the emb1507-4 allele displayed the brr2a-2 phenotype. The failure of the emb1507-4 null allele to complement the căo developmental phenotype established that altered BRR2a protein caused the căo mutant phenotype. Together with its recessive nature, these data suggest that brr2a-2 is most likely a hypomorphic and not a neomorphic allele and that the căo phenotype is most likely caused by reduced activity of BRR2a.
Threonine 895 is a strongly conserved residue

So far, no detailed data confirming the function of BRR2a proteins in Arabidopsis are available in the literature. However, it was reported that the homozygote T-DNA insertion has a drastic effect on seed viability, causing abortion at an early developmental stage (Tzafrir et al., 2004).

To better understand the critical residues and domains of this protein, yeast, animal, plant and protozoa BRR2 amino acid sequences from a region around the T895I mutation were aligned. Sequence analysis revealed that the mutated site in BRR2a-T895I (Fig. 3.7) is highly conserved among all the analyzed species.

![Figure 3.7 Threonine 895 is conserved among eukaryotic BRR2 proteins](image)

Amino acid alignment of the region where T895 is localized from several organisms. The red asterisk highlights the threonine (T) that is altered to an isoleucine (I) in brr2a-2, and the blue bar locates a conserved motif within the amino-terminal helicase cassette. Conserved amino acid residues are highlighted in black. Residues not identical but similar are highlighted in gray, dissimilar residues remain unhighlighted. Aligned amino acids were: AT5G61140 [Arabidopsis thaliana], AT1G20960 [Arabidopsis thaliana], AT1G20960 [Arabidopsis thaliana], AT1G20960 [Arabidopsis thaliana], AT1G20960 [Arabidopsis thaliana], AT1G20960 [Arabidopsis thaliana], NP_031099 [Saccharomyces cerevisiae], NP_648818 [Drosophila melanogaster], NP_00111167 [Danio rerio], NP_054733 [Homo sapiens], NP_765188 [Mus musculus], NP_00258183 [Schistosoma mansoni], NP_0296663 [Drosophila melanogaster], NP_00258183 [Drosophila melanogaster], NP_00222521 [Physcomitrella patens], NP_02222521 [Physcomitrella patens], APF98818 [Oryza sativa], NP_00222521 [Oryza sativa], NP_00222521 [Oryza sativa], XP_0022665 [Vitis vinifera], XP_0035467 [Glycine max], XP_0035714 [Brachypodium distachyon], XP_0035959 [Medicago truncatula], XP_0035959 [Medicago truncatula], XP_0035959 [Medicago truncatula], XP_0035959 [Medicago truncatula].

Furthermore, this amino acid resides close to the end of a highly conserved catalytic domain (ATP-dependent RNA helicase domain) and close to the ATP binding sites of BRR2 (Fig. 3.8) (Zhang et al., 2009). Taken into account that threonine is often phosphorylated, it is possible that it could be involved in BRR2a functional regulation. This possibility, however, will require additional molecular analyses in the future.

Furthermore, through the alignment of the different subdomains from the DEXDc and the helicase super family carboxy-termini from the Arabidopsis paralogues and the yeast orthologue, a high conservation of the amino acid positions at the binding sites for the bivalent cation (Mg$^{2+}$), ATP and nucleotide as well as the amino acid sequences of the different motifs was apparent (Fig. 3.8). Taken together, these results suggest an important structural and, possibly, functional role of T895 as well as the neighboring sequence.
Figure 3.8 BRR2 helicase domains are conserved between yeast and Arabidopsis

Schematic representation of the BRR2 helicase domains in Arabidopsis and S. cerevisiae. Each box represents the conserved motifs within the DExD domains and Helicase superfamily C terminal domain. The conserved residues from all presented sequences are highlighted in yellow. When there are differences in at least one sequence, the identical sequences are marked in blue and the alternative residue in green. Non-identical residues from all the sequences are not highlighted. The position of the ATP, nucleotide, and Mg$^{2+}$ binding sites are marked by red asterisks, blue bars and orange bars, respectively. The position of the conserved mutated threonine is highlighted by a green arrow head. Amino acid sequences from S. cerevisiae Brr2p and from different Arabidopsis paralogues were aligned using ClustalW software (Tamura et al. 2011). The position of the different domains was predicted using SMART protein prediction software (Letunic et al. 2009).

BRR2a is a highly conserved protein

To evolutionary characterize the BRR2a protein, its amino acid sequence was obtained from TAIR (http://www.arabidopsis.org) and blasted against the non-redundant protein data bases at NCBI. Representative sequences form different eukaryotes including yeast, insects, mammals, protozoa and plants were aligned using ClustalW software (Larkin et al, 2007). A neighbor-joining tree was constructed based on the multiple alignments using MEGA5 software. The phylogenetic tree showing the relationship among all the BRR2 proteins analyzed is presented in (Fig. 3.9). The tree was based on the p-distance, which is the proportion of amino acid sites at which the two sequences compared are different. To assess the robustness of the constructed Neighbor-joining tree, bootstrap values from 1000 replicates were displayed.
It is interesting to note that the BRR2a (AT1G20960) protein is evolutionarily conserved among eukaryotes from yeast to human (Fig. 3.9). The analysis of the Arabidopsis genome had revealed the presence of at least three BRR2 genes: BRR2a (AT1G20960), BRR2b (AT2G42270) and BRR2c (AT5G61140). The phylogenetic analysis had revealed that all the analyzed organisms have at least one copy of a BRR2 gene: two copies in yeast, rice and humans, and three copies in soybean. Furthermore, the analysis of the degree of homology among yeast Brr2p and the Arabidopsis paralogues (Tab. 3.3) had revealed that BRR2a shares a 80% identity and 90% similarity with BRR2b, compared to 40 and 60% respectively with BRR2c. The high identity between the BRR2a and BRR2b suggests a conserved structure and probably function. They thus may act redundantly during plant development. In contrast to this notion, mutants in BRR2a alone have strong phenotypes arguing against widespread redundancy.

Figure 3.9 Phylogeny of Brr2p homologues.
BRR2 protein sequences of several organisms were aligned and a phylogenetic tree was generated based on this alignment. Branch lengths indicate distances. Numbers on the branch are bootstrap values of confidence in the displayed branches. CAA94089.1, AAS78571.1 [Homo sapiens] (Hs); EAZ28547.1 [Oryza sativa] (Os); NP_001116729.1 [Danio rerio] (Dr); CAA97301.1, NP_011099.1 [Saccharomyces cerevisiae] (Sc); NP_001185050.1 [Arabidopsis thaliana] (BRR2a), NP_181756.1 (BRR2b), NP_200922.2 (BRR2c); NP_002581343.1 [Physcomitrella patens] (Pp); XP_002173505.1 [Schizosaccharomyces japonicus] (Sej); XP_002318725.1, XP_002322252.1 [Populus trichocarpa] (Pt); XP_002581343.1.
Table 3.3 Amino acid identity and similarity between yeast Brr2p and Arabidopsis BRR2 paralogues

<table>
<thead>
<tr>
<th>AT2G42270</th>
<th>AT5G61140</th>
<th>Brr2p</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRR2b</td>
<td>BRR2c</td>
<td></td>
</tr>
<tr>
<td>AT1G20960 BRR2a</td>
<td>82/91</td>
<td>40/59</td>
</tr>
<tr>
<td>AT2G42270 BRR2b</td>
<td>39/58</td>
<td></td>
</tr>
<tr>
<td>AT5G61140 BRR2c</td>
<td></td>
<td>30/52</td>
</tr>
</tbody>
</table>

**BRR2A is similarly expressed in Col and in brr2a-2**

To investigate the effect of the mutation in brr2a-2 on the expression level of BRR2a, the transcript level of the altered gene was investigated using quantitative RT-PCR in both Col and brr2a-2. No significant differences in the transcript levels between the wild-type Col and the mutant plants were observed (Fig. 3.10). This suggests that the mutation has no effect on the BRR2a expression level. Effects on protein function, however, cannot be excluded.

**Figure 3.10 BRR2a is expressed at similar levels in Col and brr2a-2**

Quantitative RT-PCR was performed on cDNA from Col and brr2a-2. RNA was extracted from 15 day-old seedlings grown under SD conditions, at zeitgeber (ZT) =7. Relative expression to PP2a is shown as mean ± S.E.M. (n =3).
**BRR2a is ubiquitously expressed in Arabidopsis**

In order to investigate the spatial expression level of the three Arabidopsis paralogues, the expression levels of the *BRR2a, BRR2b* and *BRR2c* genes were analyzed in silico using the Arabidopsis eFP ‘Electronic Fluorescent Pictograph’ browser online software (Winter et al., 2007); eFP browser is a tool used for the interpretation and visualization of genes expression data collected from several *A. thaliana* microarrays, absolute and relative expression values can also be obtained from this tool.

The results presented in (Fig. 3.11) suggest that the *BRR2* genes are expressed in most tissues at variable levels. *BRR2a* seems to be expressed strongest of all Arabidopsis *BRR2* genes with highest signals in dry seeds. *BRR2c* was expressed considerably weaker than *BRR2a* but had also highest signal in seeds. *BRR2b*, which is most similar to *BRR2a* in sequence, was expressed much weaker in all organs. Often, duplication of genes can be followed by a slow decrease in the expression of one of the copies, which eventually will turn into a pseudogene, and *BRR2b* and *BRR2c* may be other examples of such a situation.

![Figure 3.11 Expression profiles of the Arabidopsis BRR2 paralogues in different organs](image)

**Figure 3.11 Expression profiles of the Arabidopsis BRR2 paralogues in different organs**

The expression profile of the different *BRR2* paralogues was investigated in different organs according to the Arabidopsis eFP Browser (Winter et al., 2007). The y-axis shows the expression level presented as absolute expression values, the data source used was the developmental series.

**FLC expression levels are altered in brr2a-2**

*FLC* is a MADS box DNA binding protein and a major repressor in the control of flowering time in Arabidopsis (Michaels and Amasino, 1999). Many early flowering Arabidopsis mutants have reduced *FLC* expression while many late flowering mutants have increased *FLC* expression.

The reference ecotype Col has intermediate *FLC* expression because it lacks an active allele of the *FLC* activator *FRIGIDA* (*FRI*).

Col plants with an introgressed *FRI* allele have strongly increased *FLC* expression. To investigate possible alterations in *FLC* transcript levels in *brr2a-2*, *FLC* expression levels were measured in 15 days old Col, *brr2a-2* and *FRI* seedlings grown under SD conditions. As reported before, the expression levels of *FLC* were nearly four times higher in *FRI* than in Col (Michaels and Amasino, 2001).
In contrast, FLC levels were reduced by more than 90 times in brr2a-2 (Fig. 3.12 A). It has been reported that 3’ end processing of FLC is a sensitive process that is altered in some mutants (Liu et al, 2010). To test whether brr2a-2 has defects in 3’ end processing and polyadenylation of FLC, RT-PCR was performed on cDNA generated after enzymatic in vitro polyadenylation of isolated RNA. Using primers that amplify the full length FLC cDNA, a band could only be obtained from Col but not brr2a-2 RNA (Fig. 3.12 B).

Taken together, the FLC mRNA level was strongly reduced in brr2a-2. These results correlate with the observed early flowering phenotype, and are consistent with previous reports about the effects of FLC on flowering time control (Michaels and Amasino, 1999).

Figure 3.12 FLC expression level is altered in brr2a-2
(A) Expression of the flowering repressor gene FLC in Col, brr2a-2 and FRI. Quantitative RT-PCR was performed using RNA from 15 day-old seedlings grown under SD conditions, at ZT =7. cDNA was synthesized with oligo dT primers. Relative expression to PP2a is shown as mean ± S.E.M. (n =3). (B) RT-PCR amplification of full-length FLC from Col and brr2a-2. mRNA was isolated as described above, cDNA was synthesized after in vitro polyadenylation, and primers corresponding to the full length FLC were used for RT-PCR. GAPDH was used as a loading control

MAF gene expression is altered in brr2a-2

The MADS AFFECTING FLOWERING (MAF) gene family is a MADS box transcription factor gene family that is closely related to FLC both at the amino acid level and in exon-intron gene structure. In the Arabidopsis genome, there are five MAF genes; similarly to FLC they act as negative regulators of the floral transition (Ratcliffe et al, 2001, Ratcliffe et al, 2003, Zografos and Sung, 2012). Because FLC expression was strongly affected in brr2a-2, the levels of the MAF transcripts were also investigated. Significant reductions in expression levels were found for MAF1 and MAF4 while MAF2, MAF3 and MAF5 showed a trend of reduced expression without being statistical significant (Fig. 3.13 A, B).
Figure 3.13 MAF genes expression levels are moderately affected in brr2a-2

Expression level of MAF1 (A) and MAF2-5 (B) genes in Col and brr2a-2. Quantitative RT-PCR was performed on cDNA from Col and brr2a-2. RNA was extracted from 15 day-old seedlings grown under SD conditions at ZT=7. cDNA was synthesized with oligo dT primers. Relative expression to PP2a is shown as mean ± S.E.M. (n = 3).

These results show that the expression levels of the MAF genes family members were also affected in brr2a-2, but that the changes of FLC were much larger.

**FT and SOC1 expression is increased in brr2a-2**

FLOWERING LOCUS T (FT) and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1, AGL20), often referred to as floral integrators, are promoters of flowering that both are repressed by FLC. FT activation was reported before to be delayed in msi1-tap1 (Exner et al, 2010). To test whether the early flowering of brr2a-2 was associated with increased FT and SOC1 expression, the transcript levels of these floral activators were measured in Col and brr2a-2. Indeed, the expression levels of FT were nearly three times higher in brr2a-2 than in Col (Fig. 3.14 A). In addition, also SOC1 had significantly increased expression (Fig. 3.14 B). These results are consistent with the decreased FLC levels and suggest that early flowering in brr2a-2 is caused by increased expression of FT and SOC1 that have been freed from the repression by FLC.
Figure 3.14 *SOC1* and *FT* expression are increased in *brr2a-2*.
Expression of flowering time activator genes *FT* (A) and *SOC1* (B) in Col and *brr2a-2*. Quantitative RT-PCR was performed on cDNA from Col and *brr2a-2*. RNA was extracted from 15 day-old seedlings grown under SD conditions at ZT=7. Relative expression to *PP2a* is shown as mean ± S.E.M. (n =3).

Known *FLC* regulators are normally expressed in *brr2a-2*

*FLC* expression is tightly controlled by various activators and repressors, in particular those from the autonomous flowering pathway. Seven genes (*FCA, FPA, FY, FVE, LD, FLD, FLK*) have been identified as members of the autonomous flowering pathway, all of which affect flowering by constitutively repressing *FLC* (Michaels and Amasino, 2001, Lim et al, 2004). It was possible that strongly increased expression of one or more of the autonomous pathway genes caused reduced *FLC* expression in *brr2a-2*.

In order to test this hypothesis, the expression of *FLC* repressors was analyzed using RT-qPCR under the previously described conditions. No differences in the expression of any of the analyzed genes were observed between Col and *brr2a-2* (Fig. 3.15). This suggests that the *FLC* repressors from the autonomous pathway analyzed in this work were not responsible for the decreased *FLC* expression in *brr2a-2*. 
Figure 3.15 Expression of FLC repressors was not altered in brr2a-2. Quantitative RT-PCR was performed on cDNA from Col and brr2a-2. RNA was extracted from 15 day-old seedlings grown under SD conditions at ZT=7. Relative expression to PP2a is shown as mean ± S.E.M. (n =3).

It was also possible that decreased FLC expression in brr2a-2 was caused by decreased expression of FLC activators. To test this hypothesis, the expression of known FLC activators from the FRI-mediated pathway “FES1, FRL2, SUF4” (Schmitz et al, 2005, Choi et al, 2011), the PAF1 complex “ELF7, ELF8, VIP4, VIP5” (He et al, 2004) and RNA-processing factors that affect FLC “HUA2, SERRATE, ABH1, PEP” (Doyle et al, 2005, Bezerra et al, 2004, Ripoll et al, 2009) was tested as described above. No significant change in the expression of any of these genes was observed in brr2a-2 (Fig. 3.16). This suggests that the FLC activators analyzed in this work were not responsible for the decreased FLC expression in brr2a-2. Because known FLC activators and repressors are not likely to have caused the reduced FLC transcript levels in brr2a-2, it appeared possible that an uncharacterized mechanism had affected FLC levels.
Figure 3.16 Expression of FLC activators was not altered in brr2a-2. Quantitative RT-PCR was performed on cDNA from Col and brr2a-2. RNA was extracted from 15 day-old seedlings grown under SD conditions at ZT=7. Relative expression to PP2a is shown as mean ± S.E.M. (n =3).

Transcript processing of FLC is altered in brr2a-2

To investigate whether the altered FLC transcript levels observed in brr2a-2 plants were due to defects in FLC transcript splicing, the splicing of three introns of FLC was measured in both Col and brr2a-2. Using quantitative RT-PCR, the relative levels of spliced and intron-containing, un-spliced transcripts were determined. To quantify potential splicing defects, the unspliced transcript proportion, defined as the ratio of unspliced to total transcript (spliced + unspliced), was calculated for both Col and brr2a-2. Indeed, the unspliced transcript proportion was strongly affected and was about 8-fold higher in brr2a-2 than in Col (Fig. 3.18A, B). These results suggest that a defect in the splicing machinery caused reduced accumulation of correctly spliced FLC in brr2a-2. It is likely that incorrectly spliced transcripts are subject to the Nonsense Mediated Decay mRNA quality control and have a much faster turnover than correctly spliced transcripts (Brogna and Wen, 2009).

Splicing efficiency is affected by many parameters, and intron secondary structure is one of them (Warf and Berglund, 2010, Koornneef et al, 1998). Recently, it has been reported that budding yeast Brr2p has a particular role in splicing of highly structured introns (Hanh et al, 2012). To test whether FLC introns are particularly strongly structured, secondary structure formation was assessed for all Arabidopsis introns and expressed as the minimum free energy (MFE) as done for yeast in (Hanh et al, 2012). It became obvious that FLC introns 1 and 6 were predicted to be highly structured. FLC intron 1 was predicted to belong to the 0.5% of most structured Arabidopsis introns (data not shown). Taken together, these results show that FLC splicing is impaired in brr2a-2 mutants. It is possible that the prevalent effect on FLC was caused by a large degree of secondary structure in FLC introns.
Figure 3.18 *FLC* splicing efficiency is reduced in *brr2a-2* (A) *FLC* genomic structure, exons are depicted as black boxes, introns as black lines. Positions of regions encompassed by different primer pairs used to amplify distinct PCR products are indicated. (B) The unspliced transcript proportion was calculated as the ration of unspliced to total transcript (US/(S+US)) for three representative *FLC* introns in both Col and *brr2a-2*.

BRR2a-T895I affects the splicing efficiency of other genes

Splicing defects in *FLC* can explain the early flowering phenotype but we wondered whether *brr2a-2* affects only *FLC* or also other genes. To this aim, two genes were selected for the analysis, *MAF1* and *IBM1*, and the unspliced transcript proportion was determined for their introns I1 and I7, respectively. The unspliced transcript proportion of the intron I1 from *MAF1* was nearly two fold increased in *brr2a-2* compared to Col (Fig. 3.19 A, B). There was also a slight increase in the unspliced transcript proportion for intron 7 of *IBM1* in *brr2a-2* compared to Col (Fig. 3.19 A, C). Taken together, these results suggest that *FLC* might be strongest affected in *brr2a-2* but that also other genes have splicing defects in this mutant.
Figure 3.19 Splicing efficiency of *MAF1* and *IBM1* is reduced in *brr2a-2*

(A) *MAF1* and *IBM1* genomic structure, exons are depicted as black boxes, introns as black lines. Positions of regions encompassed by different primer pairs used to amplify distinct PCR products are indicated. (B) Unspliced transcript proportion (US/(S + US)) for intron I1 of *MAF1*. (C) Unspliced transcript proportion for intron I7 of *IBM1*. RNA was extracted from 15 day-old seedlings grown at SD conditions at ZT=7. Quantitative RT-PCR was performed on cDNA from Col and *brr2a-2*; results were normalized to *PP2a*.

In addition, we performed a gel-based RT-PCR screen for potential splicing defects in other genes using exon and intron gene specific primers. Four additional genes with different intron sizes were selected for analysis. For these genes, however, no signs of increased splicing defects could be detected in *brr2a-2* (Fig. 3.20). Although these results suggest that splicing is not grossly altered for these genes in *brr2a-2*, more sensitive quantitative assays of spliced transcript proportions may reveal some effects even on these genes.
Figure 3.20 the splicing patterns of several intron-containing transcripts in Col and brr2a-2 using RT-PCR

(A) Schematic diagram of the MAF3, MAF4, MAF5, and AGL15 gene structure, exons are depicted as black boxes, introns as black lines. Positions of regions encompassed by different primer pairs used to amplify distinct PCR products are indicated by arrows. (B) RT-PCR was performed from cDNA obtained as described above, and the PCR product was separated on a 1% agarose gel.

Taken together, the BRR2a protein, which is a key regulator of splicing in yeast and mammals, is also required for efficient splicing of at least a subset of genes in plants. If lost it causes seed abortion and the BRR2a-T895I version causes early flowering by decreasing levels of correct FLC transcripts.

DISCUSSION

In Arabidopsis, the regulation of gene expression during development is best studied at the transcriptional level. However, many reports suggest an important role also for RNA processing in the control of development and in particular the floral transition (Quesada et al, 2005). A set of genes, which is commonly referred to as the autonomous flowering pathway, control flowering time by affecting the expression and transcript processing of FLC. This set includes FCA, FY, FPA, FLK, PEP and ELF9 (Simpson et al, 2003, Schomburg et al, 2001, Lim et al, 2004, Ripoll et al, 2009, Song et al, 2009). Independently from the autonomous pathway, other genes appear to affect FLC transcript processing although their biochemical modes of action remain often unknown. For instance, HUA1, HUA2 and HEN4 promote the processing of some flowering time genes and are required for high FLC expression (Cheng et al, 2003), and the misexpression of ABA HYPERSENSITIVE1 (ABH1), which encodes the large subunit of the nuclear mRNA cap binding complex, leads to an accumulation of a partially spliced intron 1 of FLC and premature polyadenylation (Kuhn et al, 2007). In addition, a mutation in the nuclear cap binding protein CBP20 resulted in very low FLC mRNA levels and an increased ratio of unspliced to spliced FLC transcript (Geraldo et al, 2009). Mechanisms of transcript processing for FLC are still poorly understood (Lorkovic, 2009). However, genetic studies had suggested that several of the genes mentioned above act in
parallel and independent pathways. In this work, I demonstrate the function of another pre-mRNA processing protein (BRR2a), in the vegetative to reproductive developmental stage transition.

**BRR2a is involved in Arabidopsis flowering time**

Flowering time analysis clearly demonstrated the early flowering phenotype of brr2a-2 both under SD and LD photoperiods. Comparisons with brr2a-2 msi1-tap1 double mutants revealed an additive rather than epistatic genetic interaction. This suggests that MSI1 and BRR2a control flowering time through two independent pathways. The suppression of the late flowering phenotype of msi1-tap1 by brr2a-2 is possibly caused by the alteration of the expression levels of one or more of the flowering time integrator genes FT, SOC1 and FLC.

*FT* acts downstream of several flowering time pathways including the photoperiod and autonomous pathway (Amasino and Michaels, 2010, Turck et al, 2008). *FT* expression is mainly localized to the vasculature of cotyledons and leaves of plants grown in LD.

*FT* protein then migrates to the shoot apical meristem, where it binds to a bZIP transcription factor, FD, and together with FD induces the expression of floral meristem identity genes such as APETALA1 (AP1) (Adrian et al, 2010). *FT* is a very potent inducer of flowering, and not surprisingly many early flowering mutants have high *FT* expression in both SD and LD (Yu and Michaels, 2010, Sawa and Kay, 2011, Lu et al, 2011). *FT* expression is strongly reduced in msi1-tap1 (Exner et al, 2009) explaining the late flowering phenotype in LD. In contrast, *FT* has strongly increased expression in brr2a-2 explaining the early flowering phenotype in LD and SD. *SOC1* is another potent inducer of flowering and functions downstream of *FT* in LD and independently of *FT* in SD (Lee and Lee, 2010). Reduced *SOC1* expression was decreased in msi1-tap1 explaining the late flowering phenotype in SD (Bouveret et al, 2006). In contrast, *SOC1* expression was increased in brr2a-2 most likely as a consequence of increased expression of its activator *FT*. Together, decreased and increased expression of *FT* and *SOC1* can explain the late and early flowering times of msi1-tap1 and brr2a-2.

**Expression of FLC and MAFs is altered in brr2a-2**

What could be the reason for increase *FT* and *SOC1* expression in brr2a-2? One explanation is *FLC*, which prevents the flowering transition through its direct binding to *SOC1* and *FT* (Helliwell et al, 2006). In many flowering time mutants, *FLC* expression is strongly altered. Similarly, in brr2a-2 levels of mature *FLC* transcripts were very low, providing an explanation for the up-regulation of the two floral integrators *FT* and *SOC1*. Transcript levels of the *FLC* homologs MAF1-5 are often regulated by the genes that also affect *FLC* (Oh et al, 2004, Yoo et al, 2011), and *MAF1-5* transcript levels were altered in brr2a-2. However, no changes in expression of known regulators of *FLC* were found in brr2a-2. Thus, reduced *FLC* transcript levels can explain the early flowering phenotype of brr2a-2 plants but might not be caused by alterations of known *FLC* regulators.

**brr2a-2 alters FLC transcripts processing**

BRR2a is a member of U5 snRNP proteins, which in animals and yeast are stably bound to the spliceosome complex during transcript processing. In animals and yeast, BRR2 proteins play key roles in intron splicing. Here, we demonstrated that Arabidopsis BRR2a plays a similar role and that its mutation resulted in an early flowering phenotype through defects in transcript processing of some flowering time genes. Obvious splicing defects were observed on several introns, including intron I1 (=3500bp), I5 (194) and I6 (992) of *FLC*. 

55
Some defects in splicing were observed for MAF1 and IBM1 as well. Thus, introns and genes differ in their dependency on BRR2a, and FLC is particularly sensitive to BRR2a loss. Interestingly, while the proportion of intron-containing FLC transcripts increased, total amounts of FLC transcripts decreased in brr2a-2. This is similar to earlier observations in pep-4 mutants (Ripoll et al, 2009). This effect is possibly due to the action of RNA quality pathways such as the nonsense-mediated mRNA decay (NMD) pathway, a cotranscriptional quality control pathway, which degrades aberrant transcripts (Brosgnà and Wen, 2009, Gonzalez et al, 2001). In addition to the NMD pathway, other RNA quality pathways seem to operate in the nucleus and cause degradation of transcripts with delayed processing independent of the presence of stop codons (Porrua and Libri, 2013). It remains to be established, to which extend RNA quality pathways can feed back to repress transcription. Notably, a functional correlation between transcription and transcript processing was previously reported in yeast and mammals (Alexander et al, 2010).

Complete loss of Arabidopsis BRR2a is lethal (Tzafrir et al, 2004), and the relatively mild phenotype of brr2a-2 plants suggests that brr2a-2 is a hypomorphic allele and that BRR2a-T895I has partially reduced function. Nevertheless, brr2a-2 had pleiotropic morphological defects revealing that BRR2a function is not restricted to flowering control, but is also needed for other vegetative (leaves) and reproductive (siliques, seeds, flowers) developmental programs. The important role of BRR2a in development is consistent with previous reports that confirmed the essential role of spliceosome components for cell viability and proliferation in Arabidopsis and other organisms (Otake et al, 2002, König et al, 2007, Gross-Hardt et al, 2007, Volz et al, 2012, Kuhn et al, 2002, Kim et al, 2010, Liu et al, 2009). Furthermore, the generally mild phenotype changes of brr2a-2 plants suggest that the splicing of only a group of genes is affected. These results agree with previous studies. Mutation of Arabidopsis SMD3-b, a spliceosome subunit, revealed that only some genes had altered transcript abundance and changes in splicing efficiency and that these genes were responsible for most aspects of the smd3-b mutant phenotype (Swaraz et al, 2011). In skip-1 mutants, which are defective in another component of the spliceosome, changes in alternative splicing of some circadian clock and flowering time genes was observed (Wang et al, 2012). Similar results were obtained in stip1 mutants lacking another spliceosome member, in which only a subset of circadian clock and flowering time genes were affected (Jones et al, 2012).

Why do some spliceosome mutants affect only some introns or genes? There is no final answer to this question yet, but it is becoming clear that some spliceosomal proteins prefer particular splice site or intron sequences (Lorkovic et al, 2005, Brown et al, 2002, Hu et al, 2011) or are differentially affected by pre-mRNA secondary structure (Warf and Berglund, 2010).

A recent paper from the Beggs laboratory had shed some light on the function of Brr2p in the catalytic center of the yeast spliceosome. Using in vivo UV cross-linking and RNA sequencing in budding yeast, Brr2p was found to bind to the U4/U6 stem I and to the loop 1 of the U5 snRNA and near splice sites (Hahn et al, 2012). Interestingly, Brr2p displayed a preference for introns with long and structured sequences between the branching point (BP) and 3’ splice site (3’ SS). The Brr2p-G858R variant had changed binding affinities and processing of the longer, structured introns was hampered. G858 in yeast BRR2p is only 32aa away from the threonine that corresponds to T895 in Arabidopsis BRR2a, which was mutated in brr2a-2 plants. Notably, intron 11 of FLC is one of the longest and most structured introns in the Arabidopsis genome, and its processing is greatly affected in brr2a-2 mutants. Furthermore, the observed splicing defects mirrored the amount of predicted secondary structure for FLC introns, supporting the idea of BRR2a function to bind to structured introns. In yeast, it was mainly the amount of structure between BP and 3’SS that determined dependency on Brr2p.

56
Interestingly, intron 5 of FLC was among the 0.7% of Arabidopsis introns that were predicted to be most strongly structured between predicted BP and 3’SS while intron 6 of FLC was among the 0.5% of Arabidopsis introns that were predicted to be most strongly structured overall. It remains to be tested whether less efficient splicing of individual introns can feed back on the splicing of other introns in the same pre-mRNA.

BRR2a is a highly conserved spliceosomal protein

Unlike plant BRR2, yeast and human BRR2 proteins are very well characterized. Brr2p, also known as Snu246 in yeast and as U5-200kD in humans (Laggerbauer et al, 1998), is an evolutionary highly conserved spliceosomal protein of about 200 kDa. This nuclear protein (Boon et al, 2007) was first identified in yeast, when screening for genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations (Noble and Guthrie, 1996). The protein belongs to the DEAD/DEXH-box family of ATP-dependent RNA helicases with two putative RNA helicase domains, each with a highly conserved ATPase motif (Kim and Rossi, 1999), followed by a sec63 domain (Ponting, 2000). The yeast Brr2p was immunopurified in large complexes containing U1, U2, U5 and duplex U4/U6 snRNPs (Raghunathan and Guthrie (1998). The purified yeast Brr2p and its homologue in mammalian cells were reported to be components of the spliceosome throughout the splicing cycle. Its ATPase and helicase activities are involved in rearrangements necessary for spliceosome activation through the unwinding of U4/U6 snRNP (Raghunathan and Guthrie (1998). After the unwinding, Brr2p remains stably associated with the catalytic core of the spliceosome (Bessonov et al, 2008) and eventually promotes spliceosome disassembly by unwinding U2/U6 (Small et al, 2006). Brr2p functions as well in promoting conformational rearrangements in the spliceosome during the first-to-second-step transition, which aid 3’ SS positioning and formation of the second-step catalytic center (Hahn et al, 2012). Brr2’s helicase activity is highly regulated by the different U5 snRNPs to ensure the correct timing of spliceosomal activation or disassembly (Bottner et al, 2005); regulators of BRR2p function include Prp8 (Bellare et al, 2006, Mozaffari-Jovin et al, 2012) and Snu114 GTPase (Small et al, 2006).

The amino acid alignment of the different Arabidopsis paralogues with their orthologues from different eukaryotes clearly demonstrated the strong conservation of the amino acids located at helicase domains I and II. Evolutionary conservation of BRR2 proteins suggests a structural and functional conservation. Furthermore, Arabidopsis has three BRR2 genes, while yeast and humans have two genes. BRR2a is ubiquitously expressed in most plant tissues with highest transcript accumulation in seeds. This is consistent with the seed abortion phenotype at early development stages of homozygous null mutants (Tzafrir et al, 2004). BRR2b is very similar to BRR2a but BRR2b is much weakly expressed in most tissues. The third BRR2 gene in Arabidopsis, BRR2c, has more differences to BRR2a and has lower expression. Its divergence in amino acid sequences may suggest possible biochemical and functional differences. Future studies need to establish whether the Arabidopsis BRR2 proteins have shared or specific functions.

A conserved threonine at position 895 is important for BRR2a function

Threonine T895 and its neighboring sequences are highly conserved and are located near the ATPase domain within the first conserved helicase I motif (Zhang et al, 2009). A crystal structure and structural models for different yeast and human BRR2 helicase regions had revealed a possible reorganization and pairing of these domains during the splicing process (Santos et al, 2012).
Furthermore, mutagenesis studies in the amino-terminal helicase domain had revealed the important role of this domain in splicing efficiency and cell viability (Plumpton et al., 1994; Staley and Guthrie, 1998; Brenner and Guthrie, 2005). Taken together, the T895I mutation in brr2a-2 may lead to a partial loss of BRR2a’s splicing function.

Based on observations from the mammalian and yeast models, it is tempting to speculate that the amino-terminal helicase domain may play a critical role in the splicing of highly structured Arabidopsis introns. It is possible that the T895I mutation impairs interaction of BRR2a with other proteins. Plausible interactors are AT4G38780 and GAMETOPHYTE FACTOR 1 (GFA-1) (Liu et al., 2009), which are Arabidopsis homologues of the Brr2p regulatory yeast proteins Prp8 and Snu114 (Maeder et al., 2009), respectively. Indeed, the carboxy-terminal domain of BRR2a interacted in a yeast two hybrid assay with GFA1 (Liu et al., 2009). However, it is not known whether mutation of T895 in the amino-terminal domain affects this interaction. Other candidates of BRR2a interactors are homologues of the serine/threonine protein kinase Prp4 and its substrates such as Prp1. These proteins interact genetically and biochemically with fission yeast Brr2p and Prp8. These interactions take place in spliceosomal complexes and control the structural rearrangements towards the activation of spliceosomes. No evidence for the phosphorylation of Brr2p or Prp8 in vivo was reported in fission yeast (Bottner et al., 2005; Schwelndus et al., 2001); nevertheless, phosphorylation of BRR2a by the Arabidopsis Prp4 homolog LACHESIS remains possible (Gross-Hardt et al., 2007; Volz et al., 2012). Future experiments will establish whether the T895I mutation affected BRR2a’s ability to interact with other proteins.

An alternative explanation for impaired function of BRR2a-T895I is that the T895I mutation might weaken BRR2a interactions with U4, U5 or U6 RNAs or pre-mRNAs. This appears plausible because in vivo UV cross-linking and RNA sequencing have established that budding yeast Brr2p binds to the U4/U6 stem I and to the loop 1 of the U5 snRNA and to pre-mRNAs near splice sites (Hahn et al., 2012). It is most likely that the polar T895 is surface-exposed, and if a change at this residue reduces BRR2a-RNA interactions, a reduced activity of BRR2a-T895I in the splicing reaction is possible.
In summary, a mutant screen for accelerated flowering yielded the early flowering brr2a-2 mutant. Our data suggest a model (Fig. 3.21) in which BRR2a functions in the spliceosome. The missense mutation T895I leads to a partial loss of function, with reduced splicing efficiency of some genes, most importantly FLC. Reduced FLC expression allows transcription of FT and SOC1 to accelerate flowering. In the future, it will be important to establish the basis of BRR2a’s substrate specificity. It is possible, that FLC was particularly strongly affected because it has strongly structured introns. Together, this work firmly establishes the importance of undisturbed splicing for normal control of flowering time.

ACKNOWLEDGMENT

We thank Christian Beisel at D-BSSE in Basel Switzerland for Illumina sequencing. This work was supported by grants from the Swiss National Science Foundation SNF and the Swedish Science Foundation VR as well as from the Swedish University for Agricultural Sciences.

REFERENCES


Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences of the United States of America, 101, 4712-17.


Supplemental material:
Supplemental Table S1
List of gene-specific primers used for RT-PCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer ID/Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>LH0674: CCGGATGTTGACATGCTCAGT TTCTCGACCCAGCAGCTTCAAT</td>
</tr>
<tr>
<td>FLC</td>
<td>LH1420: ATTCACAAGAATAAGGTACAAAGT</td>
</tr>
</tbody>
</table>
**Supplemental Table S2**

List of gene specific primers used for splicing assay

<table>
<thead>
<tr>
<th>Gene / Primer position</th>
<th>Primer ID/Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FLC_1</strong>&lt;sup&gt;I&lt;/sup&gt; 5&lt;sup&gt;R&lt;/sup&gt;</td>
<td>LH2039: AATATGTGTGCAAGCTCGTTAAG</td>
</tr>
<tr>
<td><strong>FLC_E</strong>&lt;sup&gt;5&lt;/sup&gt; F&lt;sup&gt;+&lt;/sup&gt;</td>
<td>LH2040: AACTCATGTTGAAGCTTGTTGAG</td>
</tr>
<tr>
<td><strong>FLC_E6</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2041: CCTGGTTCTCTTTCTTTCAGCAT</td>
</tr>
<tr>
<td><strong>MAF1_E1</strong> F&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2042: CGCTGTTGTCGTCGTATCT</td>
</tr>
<tr>
<td><strong>MAF1_I1</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2043: AAATCGAGTGAGACCCAGAAA</td>
</tr>
<tr>
<td><strong>MAF1_E2</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2044: TGAGGTTTCTCCGGCTTGA</td>
</tr>
<tr>
<td><strong>IBM1_E7</strong> F&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2047: CTAGATCCACCTCGACAT</td>
</tr>
<tr>
<td><strong>IBM1_I7</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2048: AGAATGGCAGAGACTAGAAGAAAGA</td>
</tr>
<tr>
<td><strong>IBM1_E8</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH2049: TTGCATATTGGGCACTTCAGA</td>
</tr>
<tr>
<td><strong>FLC_E1</strong> F&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1801: TTCTCCACAGCTGCAACGG</td>
</tr>
<tr>
<td><strong>FLC_I1</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1802: CTCAGAAAAGTAAAAGAGCACAAAACAG</td>
</tr>
<tr>
<td><strong>FLC_E2</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1803: CACACCTTGAGACTGCCCTCTCCGT</td>
</tr>
<tr>
<td><strong>FLC_I6-2</strong> F&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1804: CACACCTTGAGACTGCCCTCTCCGT</td>
</tr>
<tr>
<td><strong>FLC_E4-1</strong> F&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1805: ACACCTTGAGACTGCCCTCTCCGT</td>
</tr>
<tr>
<td><strong>FLC_E7-1</strong> R&lt;sup&gt;-&lt;/sup&gt;</td>
<td>LH1806: ACGCTCGCCCTTTATCAGCGGAA</td>
</tr>
</tbody>
</table>

*<sup>*I</sup>: Intron, <sup>E</sup>: Exon, <sup>F</sup>: Forward primer, <sup>R</sup>: Reverse primer

**Supplemental Table S3**

List of gene-specific primers used for RT-qPCR

<table>
<thead>
<tr>
<th>Gene ID, Alias</th>
<th>Primer ID/Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G10930, PP2A</td>
<td>DK053: TCTTAATTCTGAATCGGGTTAG</td>
</tr>
<tr>
<td></td>
<td>DK054: CATACTAACAATTTACAACGCA</td>
</tr>
<tr>
<td>AT1G20960, BRR2a</td>
<td>LH1556: GCCCTTGCAAAGACGATGT</td>
</tr>
<tr>
<td></td>
<td>LH1557: AGCAGCCGAAATGATCAAGT</td>
</tr>
<tr>
<td>AT1G65480, FT</td>
<td>LH0292: ACAACTGGAACACCTTGGCAATG</td>
</tr>
</tbody>
</table>

66
<table>
<thead>
<tr>
<th>Gene Id</th>
<th>Description</th>
<th>Primer 1</th>
<th>Primer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH0293</td>
<td>ACTGTTTGCCCTGCCAAGCTGTCA</td>
<td>SUPERSOC1-F452: AGCGAGCTCAAGCAAAAGGAG</td>
<td>SUPERSOC1-R541: TGGCCAAAACCTCCGCTTCA</td>
</tr>
<tr>
<td>AT5G10140, FLC</td>
<td>LH0304: GACTGCCCTCTCCGTGACTA</td>
<td>LH0305: TTCTCAACAAGCTTTCCACATGAG</td>
<td></td>
</tr>
<tr>
<td>AT1G77080, MAF1</td>
<td>LH0781: GGCTCAGGAAAACCTCTAACA</td>
<td>LH0535: GAGTGGCAGATAATTCCGAG</td>
<td></td>
</tr>
<tr>
<td>AT5G65050, MAF2</td>
<td>LH0755: GACGTTTCTTGTTATAGAAGGTG</td>
<td>LH0827: TTACTTGGACAGCGGAAGGATC</td>
<td></td>
</tr>
<tr>
<td>AT5G65060, MAF3</td>
<td>LH0541: GAGGCTTTGTGAAGACCCATC</td>
<td>LH0542: TCCCCCACTTTTATAAGGCTGG</td>
<td></td>
</tr>
<tr>
<td>AT5G65070, MAF4</td>
<td>LH0304: GACTGCCCTGCTCCGTGACTA</td>
<td>LH0305: TTCTCAACAAGCTTTCCACATGAG</td>
<td></td>
</tr>
<tr>
<td>AT5G65080, MAF5</td>
<td>LH0304: GACTGCCCTGCTCCGTGACTA</td>
<td>LH0305: TTCTCAACAAGCTTTCCACATGAG</td>
<td></td>
</tr>
<tr>
<td>AT4G16280, FCA δ γ</td>
<td>LH1747: GAACGTGGACAGCAGCAAAAGGCTGTG</td>
<td>LH1748: AGGGTGGCCTATGCGTTCTCTGG</td>
<td></td>
</tr>
<tr>
<td>AT2G19520, MSI4</td>
<td>LH1700: TGGGTTGTCATGAGAGGTCAT</td>
<td>LH1701: CAATTGACGACCCAGTGCCAG</td>
<td></td>
</tr>
<tr>
<td>AT2G43410, FPA</td>
<td>LH1702: CAACAAATCTCTGGGAAGAGTGG</td>
<td>LH1703: GAGTGGGATAGATCGTGGTTG</td>
<td></td>
</tr>
<tr>
<td>AT3G10390, FLD</td>
<td>LH1704: GATGCGCTGATCAATGGGAATG</td>
<td>LH1705: TGGCTCTTGTTGGAAGGTTG</td>
<td></td>
</tr>
<tr>
<td>AT5G13480, FY</td>
<td>LH1706: CCACCAATGTGTGGTCAGCTGGGA</td>
<td>LH1707: ATCCCGTAGCTATCTAAG</td>
<td></td>
</tr>
<tr>
<td>AT3G48430, REF6</td>
<td>LH1708: CGGAATACCGTGTTGCAATTG</td>
<td>LH1709: CCGGATAACCTTCGTCAGT</td>
<td></td>
</tr>
<tr>
<td>AT4G15880, ESD4</td>
<td>LH1710: GCTCTGGAAAATATATGGGATGGAAG</td>
<td>LH1711: GAAATGATCCATGTGGTCCGCTGAA</td>
<td></td>
</tr>
<tr>
<td>AT5G16320, FRL1</td>
<td>LH1712: TCACAAGTTCCAGTTCTCGAGCAT</td>
<td>LH1713: TTGAGGCTGCAACACATCCCCGCAG</td>
<td></td>
</tr>
<tr>
<td>AT5G16320, HUA2</td>
<td>LH1714: CTTACCACTTCCCACCTCAACGTGAT</td>
<td>LH1715: ATCTGGGGAGCAGATGACCTGGAAT</td>
<td></td>
</tr>
<tr>
<td>AT5G61150, VIP4</td>
<td>LH1716: GCCAGACCGTCAAGCGCTCAATGGGAGA</td>
<td>LH1717: ATCGCTCTGCCCTCTCAGCAG</td>
<td></td>
</tr>
<tr>
<td>AT3G07610, IBM1</td>
<td>LH2045: CCCGAGAATGTGTGTTGACAG</td>
<td>LH2046: CACCGCTCCAAATCTCATTG</td>
<td></td>
</tr>
<tr>
<td>AT2G13540, ABH1</td>
<td>LH1777: GGGCTCCAGGATCCAACTAAAGTG</td>
<td>LH1778: CTCTCATGTGGCATATCGCTGCT</td>
<td></td>
</tr>
<tr>
<td>AT3G33520, ARP6</td>
<td>LH1779: GTACCTGAGAGTGTTTATCAGCGCAG</td>
<td>LH1780: CTCCATAACACTTAGTGTGGGTTCCTC</td>
<td></td>
</tr>
<tr>
<td>AT5G62640, ELF5</td>
<td>LH1783: CACACCTGAGAGCTTCTACACCATCAGTGGT</td>
<td>LH1784: GTGCGGCAAGAGCTTCTACCATGTCCTCTA</td>
<td></td>
</tr>
<tr>
<td>Gene Symbol 1</td>
<td>Gene Symbol 2</td>
<td>Sequence 1</td>
<td>Sequence 2</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>AT1G79730</td>
<td>ELF7</td>
<td>LH1785: CTCGGACATACCTGCTGATTATTTGAC</td>
<td>LH1786: ACCAACCTAGAAGAGTAAAACCCCTGAG</td>
</tr>
<tr>
<td>AT2G06210</td>
<td>ELF8</td>
<td>LH1787: GAGGACGATGAAGAAGAGCTGCACTA</td>
<td>LH1788: CGCCTACACCCGAAGTACGGTACCTCAT</td>
</tr>
<tr>
<td>AT1G31814</td>
<td>FRL2</td>
<td>LH1789: TACTTCTGGTGGTGTGCTGCCCCAT</td>
<td>LH1790: TAGAGGACATAGACCTGTTCCGAGCAG</td>
</tr>
<tr>
<td>AT3G12810</td>
<td>PIE1</td>
<td>LH1791: TGCTGAGAGAACCAGTACAGAAAGG</td>
<td>LH1792: GAGATCGGATCATTTGGCGCTAACCC</td>
</tr>
<tr>
<td>AT2G27100</td>
<td>SERRATE</td>
<td>LH1793: CGGATATAAAACCCACACATGCTGCC</td>
<td>LH1794: AGCATCTAGGTCCTGGAGCTGCGCA</td>
</tr>
<tr>
<td>AT1G30970</td>
<td>SUF4</td>
<td>LH1795: ATTCTTATGGCGTCGCCCCAAACACT</td>
<td>LH1796: CTGCATTATCGAGTTCATCGCTGCGCA</td>
</tr>
<tr>
<td>AT4G26000</td>
<td>PEP</td>
<td>LH1799: TCTGGGTAATCAGTCTGTATTTT</td>
<td>LH1800: CTCATCCTGTGGCGCATAAAAC</td>
</tr>
</tbody>
</table>
H3K36ac is an evolutionary conserved active histone mark in plants

Reference:

Manuscript in preparation

Authorship:

I performed most of the molecular and biochemical experiments and wrote the manuscript.
H3K36ac is an evolutionary conserved active histone mark in Plants

Mahrez, Walid\textsuperscript{1,2}, Huan, Shu\textsuperscript{1}, Nanni, Paolo\textsuperscript{3}, Gruissem, Wilhelm\textsuperscript{1} and Hennig, Lars\textsuperscript{1,2,4}

Affiliations

\textsuperscript{1} Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, CH-8092, Zurich, Switzerland.
\textsuperscript{2} Swedish University of Agricultural Sciences, Department of Plant Biology and Forest Genetics and Linnean Center for Plant Biology, PO-Box 7080, SE-75007 Uppsala, Sweden.
\textsuperscript{3} Functional Genomics Center Zurich, University of Zurich/ETH Zurich, CH-8057 Zurich, Switzerland
\textsuperscript{4} Science for Life Laboratory, SE- 75105 Uppsala, Sweden

SUMMARY

In eukaryotic cells, DNA is tightly packed in the nucleus in a structure called chromatin. The basic unit of the chromatin is the nucleosome consisting of 147 nucleotides of DNA wrapped around an octamer of four different histone proteins. Histones are subject to a large number of distinct post-translational modifications, whose sequential or combinatorial action constitutes a code that defines potential transcription states. Here, we report the identification of acetylation at lysine 36 at histone H3 (H3K36ac) as an uncharacterized histone modification in \textit{Arabidopsis thaliana} and its close relative \textit{Brassica oleracea} (cauliflower). H3K36ac was found to be an evolutionary conserved modification in seed plants. H3K36ac is highly enriched in euchromatin and very low in chromocenters. Genome-wide ChIP-seq experiments revealed that H3K36ac is generally found at the 5' end of transcriptionally active genes. Independently of gene length H3K36ac covers about 500 bp immediately downstream of the transcriptional start, where it overlapped with H3K4me3 and the H2A.Z histone variant. The histone acetyl transferase GCN5 and the histone deacetylase HDA19 are required for normal steady state levels of H3K36ac in plants. Finally, we discuss potential functions of H3K36ac.
INTRODUCTION

DNA of eukaryotes is highly packed in the nucleus in a structure named chromatin. The basic unit of chromatin is the nucleosome; consisting of 147 bp of DNA wrapped around an octamer formed by two copies each of four histone proteins (H2A, H2B, H3, H4, H1). Chromatin is dynamically regulated both temporally and spatially within the nucleus. Heritable and reversible covalent modifications at the amino-terminal tails as well as in the core of the histone proteins play an important role in the regulation of gene expression through affecting chromatin structure via the recruitment of chromatin remodelers and modifiers; the latter establishes a set of chemical marks on the histone proteins. Many histone modifications have been identified, including acetylation, methylation, phosphorylation, ubiquitylation and others (Zhang et al, 2007, Bergmuller et al, 2007). It has been suggested that the established modifications act in a combinatorial manner, forming a code, named the “Histone code”, which establishes the chromatin state, regulates gene expression and thus plays an important role in development (Jenuwein and Allis, 2001, Strahl and Allis, 2000).

Using high throughput methods with improved sensitivity, specificity and resolution such as mass spectrometry, ChIP-chip and ChIP-Seq, the exact localization and combinatorial patterns of histone modifications at an individual nucleosome level can now be determined (Liu et al, 2005, Kharchenko et al, 2011). Often, such patterns are predictive of the transcription level of the underlying gene (Wang et al, 2008).

Histone acetylation is a dynamic and reversible chemical process, which involves the transfer of an acetyl group from acetyl-CoA to histone lysine residues. The acetylation of a set of conserved lysine residues neutralizes positive charges of the histone tails and may decrease their affinity for negatively charged DNA (Hong et al, 1993). Once established, acetylated lysines can be recognized by histone code readers such as Bromo domain proteins (Filippakopoulos and Knapp, 2012), which then promote the recruitment of additional ATP-dependent chromatin remodelers, and chromatin modifier complexes, which establish a relaxed chromatin state thus facilitating the recruitment of RNA polymerases and leading to the gene expression (Bannister and Kouzarides, 2011).

Histone lysine acetylation, often associated with active transcription, is known to play important roles in numerous developmental and biological processes, including the regulation of plant cell cycle, flowering time, response to environmental conditions and hormone signals (Chen and Tian, 2007).

Several amino-terminal lysine residues of histone H3 (K9, K14, K18, K23, and K27) and H4 (K5, K8, K12, K16, and K20) were reported to be acetylated in Arabidopsis (Berr et al, 2011). Histone acetylation is dynamically regulated with a usual half-life of 2-3 minutes that rarely exceeds 30-40 minutes. This rapid turn-over strongly suggests that instead of influencing long-term epigenetic memory, acetylation participates in the short-term regulation of dynamic processes on the chromatin template (Barth and Imhof, 2010). The homeostasis of the acetylation on histone lysines is maintained by histone acetyl transferases (HATs), which catalyze the histone acetylation, and histone deacetylases (HDACs), which remove the acetylation mark (Pandey et al, 2002).

Histone lysine methylation is another important epigenetic modification with both activating and repressing roles in gene expression. Histone lysine residues can be mono-, di- or trimethylated, where each distinct methylation state can confer different biological meanings depending on the methylated histone residue, the degree of methylation and the chromatin context (Liu et al, 2010a). Histone lysine methylation is catalyzed by SET domain-containing proteins and can be removed by two classes of histone demethylases: FAD-dependent monoamine oxidases and Jumonji domain (JMJ) demethylase proteins
Once established, methylated lysines constitute landmarks for binding of chromo, MBT, Tudor, W40 domain or PHD fingers containing proteins, which subsequently recruit additional protein complexes leading to the compaction or relaxation of chromatin structure (Greer and Shi, 2012).

Histone H3K36me2 and H3K36me3 are extensively studied modifications in a wide range of organisms. In yeast and mammals, H3K36me2 is localized along the transcriptionally active gene bodies, and H3K36me3 is found to peak at the 3’ end of transcribed genes (Xu et al, 2008, Roudier et al, 2011). In contrast, H3K36me3 and H3K36me2 distribution is very different in plants (Oh et al, 2008b). Genome wide studies in Arabidopsis had defined the chromatin state 1 as highly enriched in active histone marks including H3K4me3, H3K9me3 and H3K36me3 (Roudier et al, 2011).

In yeast and fungi, the mono-, di-, and trimethylation of H3K36 is catalyzed by the SET2 protein, and in mammals and Drosophila, H3K36 can be methylated by the histone methyltransferase ASH1 (Liu et al, 2010a). In Arabidopsis, H3K36me2 and H3K36me3 are both established by the SET domain-containing protein SDG8 (Xu et al, 2008).

In animals, KDM6/JMJD3 and KDM2/JHDM1 group proteins demethylate H3K36me3 and H3K36me2/1, respectively. However, no direct homologues to these proteins were identified in Arabidopsis and rice (Xu et al, 2008).

Once established in yeast and animals, H3K36me3 is found, in addition to its role in transcriptional elongation, to be recognized by the histone deacetylase complex (HDAC) RPD3S through the chromo domain of Eaf3 (Joshi and Struhl, 2005). Subsequently, histone H3 and H4 get de-acetylated, thus preventing inappropriate nucleosome displacement at the 3’ end of genes and maintains the dormancy of cryptic promoters in the coding regions, thus preventing production of potentially deleterious transcripts such as antisense transcripts (Carrozza et al, 2005). However, in plants, H3K36me2 and not H3K36me3 is associated with transcriptional elongation and cryptic promoter silencing (Oh et al, 2008b).

Unlike in yeast and animals, the histone modification sites are much less well-known in plants. In this study we have used high performance liquid chromatography (HPLC) histone separation in combination with mass spectrometry for a survey of histone modifications on the amino-terminal tail of histone H3 to test for the presence of uncharacterized histone modification sites in Arabidopsis thaliana and its close relative Brassica oleracea (cauliflower). I describe the identification and characterization of H3K36 acetylation (H3K36ac) as a novel histone modification in plants. GCN5, a HAT, and HDA19, a HDAC, contribute to maintain H3K36ac homeostasis. H3K36ac was found to be localized in euchromatic regions and excluded from the heterochromatin-enriched chromocenters. In addition, genome-wide ChIP-seq experiments revealed that H3K36ac is generally found at the 5’ end of transcriptionally active genes, where it overlaps with H3K4me3 and the H2A.Z histone variant.

MATERIALS AND METHODS

Plant material and growth conditions

The Arabidopsis thaliana wild type, T-DNA insertion, and RNAi lines were in the Columbia (Col-0) background. The following lines were previously described: HDA19-RNAi1 (Zhou et al, 2005), gcn5-1 (Benhamed et al, 2006), 35S:Flag-HDA6 (Earley et al. 2006), hda6-1 (Aufsatz et al, 2002) and sdg8-2 (Xu et al, 2008). Seeds were sown on 0.5× basal salts Murashige and Skoog (MS) medium (Duchefa, Haarlem, The Netherlands), stratified at 4°C for one day and then transferred to growth chambers for germination at 20°C under long day (LD, 16 h light) photoperiods for ten days. The seedlings were transferred into pots
containing soil and left to grow in growth chambers under (LD) conditions. Tobacco (*Nicotiana tabacum*), wheat (*Triticum aestivum*) and rice (*Oryza sativa*) were grown in a greenhouse at 22°C in LD. For rice, special watering conditions were applied. Cauliflower was obtained from a local supermarket; Norway spruce (*Picea abies*) needles were collected from an approximately 15-years-old wild-grown tree in a forest close to Uppsala (Sweden) located at 59.81° N, 17.66° E.

**RNA isolation and RT-qPCR**

RNA extraction and reverse transcription were performed as described previously (Alexandre et al., 2009) with minor modifications. The DNA-free RNA was reverse-transcribed using a RevertAid First Strand cDNA Synthesis Kit (Fermentas, Helsingborg, Sweden) according to manufacturer’s recommendations. Aliquots of the generated cDNA were used as template for PCR with gene-specific primers (Supplemental Tab. S1). qPCR was performed using SYBR green (Fermentas) on an IQ5 multicolor Real time PCR cycler (BIO-RAD, PA, USA). qPCR reactions were performed in triplicate, gene expression levels were normalized to a *PP2A* control gene, and results were analyzed as described (Simon, 2003).

**Cross-linked Chromatin Immunoprecipitation**

Cross-linked chromatin immunoprecipitation (X-ChIP) assays were performed as described (Exner et al., 2009). Approximately 1.5 g of ten-days-old cross-linked whole seedlings were used. Antibodies used for ChIP were [anti-H3 antibody #07-690 (Upstate/Millipore); anti-H3K36me3 # pAb-058-050 (Diagenode, Belgium); anti-H3K36ac #39379 (ActiveMotif); non-immune IgG #I5006 (Sigma-Aldrich)]. Quantitative real-time PCR (qPCR) using SYBR green (Fermentas, Helsingborg, Sweden) and gene-specific primers (Supplemental Tab. S2) was performed on an IQ5 multicolor Real time PCR cycler (BIO-RAD, PA, USA). qPCR was performed with three technical replicates, and results were normalized to input and presented as the enrichment level of the given modification over the level of histone H3.

**Native ChIP-seq**

Native Chromatin Immunoprecipitation (N-ChIP) was performed by Dr. Shu Huan, following a slightly modified procedure by (Bernatavichute et al., 2008). Leaf number six of plants grown for 35 days under short day conditions (8h light) was harvested at zeitgeber time ZT=7 and used for nuclei isolation. Isolated nuclei were digested with Micrococal Nuclease (New England BioLabs, NEB, Massachusetts). The supernatant, which contained mono- and dinucleosomes, was precleared using non-immune rabbit IgG (see below) and Dynabeads Protein A (Invitrogen, Stockholm, Sweden). One tenth of the precleared chromatin preparation was kept as Input control, and ¼ was used for each immunoprecipitation with 2.5 μg antibody (monoclonal anti-histone antibody #MAB3422 (Upstate/Millipore), anti-H3K36ac #39379 (ActiveMotif) and non-immune IgG #I5006 (Sigma-Aldrich)) and collected with Dynabeads Protein A (Invitrogen). After washing, DNA was extracted by phenol-chloroform extraction and ethanol/salt precipitation (Sambrook, et al. 1989). ChIP was performed in three technical replicates. The eluted DNA was sequenced at an Illumina HiSeq2000 sequencer at the Next Generation Sequencing Unit in the Functional Genomic Center (FGCZ) Zurich, Switzerland.

**Bioinformatics and ChIP-seq data analysis**

Illumina reads were mapped to the Arabidopsis reference TAIR10 genome using bowtie2 (version 2.1) (Langmead and Salzberg, 2012). SAM file output from bowtie was converted to BAM format using SAMTOOLS (version 1.4) (Li et al, 2009) and imported into R (version 2.15.2) (http://www.R-project.org/) using functions from the Rsamtools package (Carlson et al, 2012). All subsequent analysis was performed in R. Identical reads present more than 25 times were considered as PCR artifacts and
filtered out using the filterDuplReads function from package HtSeqTools (Planet et al, 2012). After normalization to identical sequencing depth, replicates were averaged and H3K36ac signals were normalized to histone signals to control for variable nucleosome density. H3K4me3, H3K36me2, H3K36me3 and H2A.Z data were taken from the literature (Zilberman et al, 2008, Oh et al, 2008a). Gene expression data are from (Shu et al, 2012)

Constructs and recombinant protein expression and purification

The coding region for the amino-terminal tail of H3 (amino acids 1-50) was amplified by PCR from Arabidopsis cDNA using primers LH1625 (CTATTCTCTTAAAGCAACAGTTCC, forward) and LH1626 (CACCATGGCTCGTACCAA, reverse). The PCR product was cloned into the pET100-D-TOPO vector according to the manufacturer’s recommendations (Invitrogen). pGEX4T1 was obtained from (Amersham Life Science, Stockholm, Sweden); pQE30-GCN5 was previously described (Benhamed et al, 2006). Recombinant amino-terminal 6xHis tagged H3(1-50) was expressed in BL21pLys Rosetta and purified on TALON metal affinity resin (Clontech, Stockholm, Sweden) according to manufacturer’s recommendations. Recombinant GST and GST-GCN5 were expressed in BL21pLys Rosetta and purified with Glutathione Sepharose 4B resin according to the manufacturer’s recommendations (Amersham Life Science). The purified proteins were stored at -20°C until used for HAT activity assays.

Histone acetyl transferase activity assays

HAT assays using recombinant GST-GCN5 expressed in E. coli were performed as follows: 1 µg of purified His-tagged histone H3 tail (1-50) was mixed with the purified recombinant GST-GCN5 or with GST alone in HAT buffer [0.25 µg/µl Acetyl-CoA (Sigma-Aldrich), 50 mM Tris-HCl (pH 8.0), 10% glycerol (v/v), 1 mM dithiothreitol, complete EDTA-free protease inhibitor (Roche, Indianapolis, IN, USA), and 10 mM sodium butyrate]. Acetylation reactions were incubated for 60 min at 30°C with gentle shaking and terminated by addition of 2× SDS sample buffer followed by heating for 5 min to 100°C. The reaction products were electrophoretically resolved on 15% SDS-PAGE gels, and subjected to protein immunoblotting using anti-H3K36ac (#39379, Active Motif) and anti-H3 antibodies (# 07-690, Upstate/Millipore).

Immunostaining

Immunostaining was performed according to (Jasencakova et al, 2000) with some modifications. Paraformaldehyde prefixed Arabidopsis root tips were washed in MTSB buffer (50 mM PIPES, 5 mM MgSO₄, and 5 mM EGTA, pH 7.9) and digested for 10 min at 37°C with a PCP enzyme mixture [2.5% pectinase, 2.5% cellulase Onozuka R-10, and 2.5% Pectolyase Y-23 (w/v) dissolved in MTSB]. After washing with MTSB, root tips were squashed on a glass slide using a cover slip, and immediately frozen in liquid nitrogen. The cover slip was gently removed, and the nuclei were covered with blocking solution (MTSB containing 3% BSA) for 1 hour at 4°C.

Next, the samples were incubated for 1 h at 4°C with anti-H3K36ac (#39379, Active Motif) primary antibodies in MTSB (containing 1% BSA and 0.1% Tween 20) and kept in a humid chamber overnight. After 3 times washing with MTSB for 5 min each, the slides were covered for 2 h at 4°C with Rhodamine-conjugated anti-rabbit IgG antibody (Thermo Scientific) diluted in MTSB containing 1% BSA and 0.1% Tween 20. The slides were gently washed with MTSB, and the DNA was counterstained with 1 µg/ml of DAPI in mounting medium (Vectashield, Vector Labs). Fluorescence signal detection and documentation was performed with a Leica DMI 4000 microscope. Images were merged and processed using Photoshop and ImageJ.
Histone extraction and purification form plants
To extract histones from inflorescences of cauliflower and Arabidopsis or from fully developed leaves of Arabidopsis, wheat, rice and spruce, approximately 3 g of material were ground to a fine powder and homogenized for 15 min in histone extraction buffer [0.25 M sucrose, 1 mM CaCl₂, 15 mM NaCl, 60 mM KCl, 5 mM MgCl₂, 15 mM PIPES, pH 7.0, 0.5% Triton X-100 including protease inhibitors cocktail (Roche) and 10 mM sodium butyrate]. After centrifugation for 20 min at 4°C and 4500 × g, pellets were dissolved in 0.1 M H₂SO₄. After further centrifugation for 10 min at 17000 × g, total histones were precipitated from the supernatant with concentrated Trichloroacetic acid to a final concentration of 33%. The histone pellet was washed twice with acetone-0.1% HCl, air dried, dissolved either in 1× Laemmli buffer and subjected to SDS-PAGE, followed by protein immunoblotting using anti-H3 antibody antibodies (#.07-690, Upstate/Millipore); and anti-H3K36ac (#39379, ActiveMotif) antibodies, or fractionated by RP-HPLC.

Dot blotting
For Dot blotting, 1 µg of acid extracted histones dissolved in water was blotted onto a Nitrocellulose Hybond ECL™ membrane (Amersham Life Science, Stockholm, Sweden) and left to air-dry for approximately 1 h. The membrane was then blocked with 5% nonfat milk in TBST, and probed with anti-H3K36ac (#39379, ActiveMotif) antibodies as described in above.

Histone fractionation by RP-HPLC and PTM identification by tandem mass spectrometry
Bulk histone extract from cauliflower was prepared as described above. The pellet was dissolved in 100 µl of water, sonicated for 1 min and centrifuged for 1 min at 10000×g. The supernatant was gently aspirated by pipetting and loaded on an ECLIPS XDB-C8 (4.6×150 mm) (Agilent) connected to an Agilent HP1100 binary HPLC system. Histone variants were separated and eluted according to (Tamura et al, 2011). The fractions containing H3 were combined, dried and dissolved in 50 mM ammonium bicarbonate pH 8.0. After reduction and alkylation, histones H3 were digested for 2 h at 37°C with Chymotrypsin (enzyme:substrate ratio 1:50) (Promega, Switzerland) or with ArgC (1:50) (Roche, Switzerland). Samples were concentrated and desalted with C18 ZipTips (Millipore, Switzerland) following the manufacturer’s recommendations. Eluted peptides in 10µl of 50% acetonitrile and 0.1% trifluoroacetic acid were injected into an LTQ-Orbitrap mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) interfaced with a nano electrospray ion source, at the Proteomics unit of the Functional Genomic Center (FGCZ) Zurich, Switzerland. Resulting MS/MS spectra were searched with MASCOT (Matrix Science, London, UK) against the Arabidopsis TAIR9 protein database. Modifications used for searches included carbamidomethyl (cysteine (C), fixed, non-histone searches); oxidation (methionine (M), variable histone and non-histone modification); acetyl (N-term protein and Lysine (K), variable, histone searches); mono-, di-, and trimethylation (arginine (R) and lysine (K), variable, histone searches). Only strictly chymotryptic or ArgC digested peptides with a maximum of 1 missed cleavage site were allowed in database searches. The monoisotopic masses of +1, +2 and +3 charged peptides were searched with a peptide tolerance of 6 ppm and an MS/MS tolerance of 0.5 Da for fragment ions. The presence of modification sites was validated by manual inspection of spectra.

Sequence alignment of H3 amino terminal tails
Amino acid sequences of H3 from representative eukaryotes were selected using PSI-BLAST searches. Sequences were aligned using the ClustalW multiple sequence alignment program implemented in MEGA5 (Garcia et al, 2007). Visualization of the amino acid alignment was performed with GenDoc; the amino acid sequences in the FASTA format are available in Supplemental Tab. S3.
RESULTS

A conserved histone H3 modification profile in Brassicaceae

We performed a survey of posttranslational modifications on histone H3 using cauliflower, a close relative of Arabidopsis. Crude histone extract from cauliflower heads was subjected to reverse phase HPLC, and the fractions corresponding to histone H3 as evident from an immunoblot were collected, pooled and digested either with Chymotrypsin or with ArgC (Fig. 3.1).

Figure 3.1. Histone separation by reversed-phase high pressure liquid chromatography (RP-HPLC). Acid extracted histones were dried, dissolved in water, and separated on a RP-HPLC column; separation gradient (Green), eluted fractions detected by UV absorbance at 280nm (Red), eluted fractions detected by UV absorbance at 214nm (Blue). The fractions from the selected peak were separated by SDS-PAGE, followed by immunoblotting using anti H3 antibodies to confirm the presence of H3 protein.
The digested peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an LTQ-Orbitrap XL instrument equipped with a CID/ETD (collision-induced fragmentation /Electron Transfer Dissociation) fragmentation to analyze longer peptides. This instrument with its high sensitivity and accuracy can distinguish between trimethylated and acetylated peptides ($\Delta m = 0.0364$ Da). The resulting raw data were submitted to MASCOT searches for post-translational modifications with a focus on acetylation and methylation.

This led to the establishment of a map of acetylation and methylation at H3 for cauliflower (Table 3.1). All the modifications listed were identified at least once in three independent experiments using ArgC, Chymotrypsin or semi-trypsin digestion.


<table>
<thead>
<tr>
<th>Type of modification/ combination</th>
<th>Reported in Animals(^1,3/4)/Arabidopsis(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K14 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K18 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K23 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K14 ACETYL-18 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K18 ACETYL -23 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K9 Mono, Di METHYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K9 Mono METHYL- K14 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K9 DIMETHYL- K14 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K 9 ACETYL-14 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K4 Mono METHYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K27 ACETYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K27 Mono, Di, Tri METHYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K36 Mono, Di, Tri METHYL</td>
<td>1, 2</td>
</tr>
<tr>
<td>K4 Mono METHYL, K14 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K4 Di METHYL, K14-18 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K4 ACETYL - K9 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K 9-14 -18-23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Di METHYL - K18 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Di METHYL- K14, 23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 DI METHYL- K23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Di METHYL - K14-18-23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Di METHYL - K14, 18 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Mono METHYL - K14-18-23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Mono METHYL - K14-23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Mono METHYL - K14-18 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Mono METHYL - K18 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K9 Mono METHYL - K14Mono METHYL - K18ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K14 ACETYL - 23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K14-18-23 ACETYL</td>
<td>--</td>
</tr>
<tr>
<td>K(18 - 23) ACETYL - K27 Mono METHYL</td>
<td>--</td>
</tr>
<tr>
<td>K18 ACETYL - K23 METHYL</td>
<td>--</td>
</tr>
<tr>
<td>K23 Mono METHYL</td>
<td>1, 4</td>
</tr>
<tr>
<td>K18 ACETYL - K27 METHYL</td>
<td>--</td>
</tr>
<tr>
<td>K18 ACETYL - K27Di METHYL</td>
<td>--</td>
</tr>
</tbody>
</table>
Most of the modifications and their combinations identified in cauliflower were previously found in Arabidopsis (Johnson et al, 2004, Zhang et al, 2007). Taken together this suggests a high conservation of the histone code between Arabidopsis and cauliflower.

**H3K36ac is a new modification in Arabidopsis.**

Interestingly, our survey identified H3K36ac as a novel H3 modification in plants (Tab. 3.1). H3K36ac was identified from two independent experiments using different enzymes (Chymotrypsin and ArgC) with a confidence localization probability of > 90%. The MS/MS spectrum of the ArgC digested histone H3 shown in Figure 3.2 details the C and Z fragmentation ions for the parent peptide with an observed mass of \( m/z = 1504.8468 \). The high mass accuracy of the LTQ-OrbitrapXL mass spectrometer can easily distinguish between trimethylation and acetylation on histone peptides (\( \Delta m = 0.0364 \text{ Da} \)). The accurate mass of the parent ion recorded was found to be consistent with acetylation (\( \Delta M = 42.0106 \text{ Da} \)) on this peptide and not trimethylation (\( \Delta M = 42.0470 \text{ Da} \)). Importantly, we detected H3K36 mono-, di-, and trimethylation on other H3 peptides, confirming that this residue can be methylated or acetylated.
Figure 3.2. Identification of H3K36ac from cauliflower. MS/MS fragmentation spectrum of the [M+2H]^2+ parent ion at m/z 1504.8474. This peptide was identified as the H3 peptide (histone H3, amino acids 27–40) derived from ArgC digested RP-HPLC-purified cauliflower H3. The position of acetylation at Lysine 36 is displayed; the scale for the y-axis represents the relative abundance of the parent ion. Above the spectrum is the peptide sequence in which predicted c-type ions, which contain the N terminus of the peptide, are immediately above the sequence. Predicted z-type ions, which contain the C terminus, are immediately below the sequence; ions observed in the spectrum, which represent masses associated with the fragmented peptides from the MS/MS analyses, are underlined.

To validate the mass spectrometry results, histone extracts from Arabidopsis inflorescences and cauliflower seedlings and inflorescences were probed with a polyclonal anti-H3K36ac antibody on dot blots. The used antibody was previously tested for specificity using peptide array (Egelhofer et al, 2011). Positive signals were obtained from all analyzed tissues and organisms, suggesting high conservation of this modification (Fig. 3.3A).

To control for potential non-specific interactions of the antibody with other plant proteins, protein immunoblotting was performed with histone extracts from Arabidopsis inflorescences using anti-H3K36ac antibody. A single band at the size of histone H3 was obtained, confirming the presence of H3K36ac in Arabidopsis (Fig. 3.3B).
Figure 3.3. H3K36ac is present in both Arabidopsis (inflorescences) and cauliflower (seedlings, inflorescences). A) Dot-blot analysis of 1 µg of acid extracted histones from Arabidopsis (Arabidopsis thaliana, At) inflorescence, and cauliflower (Brassica oleracea, Bo) (inflorescence/ seedlings) using anti-H3K36ac antibodies. B) Acid-extracted histones prepared from Arabidopsis inflorescences were resolved on a 15% SDS-PAGE gel, followed by protein immunoblotting using anti-H3 and anti-H3K36ac antibodies.

H3K36ac is enriched in euchromatic regions of the nucleus

To establish the localization of H3K36ac in the nucleus, paraformaldehyde-fixed Arabidopsis root-tip nuclei were immuno-stained using anti-H3K36ac antibodies. H3K36ac fluorescence signals were absent from nucleoli and DAPI-dense chromocenters, which are comprised of pericentromeric repeats and other heterochromatic sequences. In contrast, H3K36ac staining was observed in the euchromatic regions (Fig. 3.4). This observation suggests that H3K36ac is mainly localized on less condensed, gene-rich regions.
Figure 3.4. Localization of H3K36ac in root tip nuclei. Root tip interphase nuclei of Arabidopsis plants were stained with DAPI and analyzed for immunofluorescence with anti-H3K36ac antibodies. A) DAPI, B) H3K36ac, C) overlay. A section through the nucleus was analyzed using ImageJ software. D) Quantitative line profiles of DAPI (blue), and H3K36ac (red) fluorescence intensities along the white line shown in (C). Scale bars: 5 µm

H3K36ac is present on active genes

To investigate the localization of H3K36ac in chromatin, X-ChIP assays were performed. Chromatin from 11 days old, cross-linked Arabidopsis seedlings was subjected to immuno-precipitation using anti-H3K36ac and anti-H3 antibodies. Precipitated DNA was probed using primers targeting an active (ACTIN 7) and an inactive (Transposon Ta2) gene (Fig. 3.5).

Figure 3.5. ChIP analysis of distribution of H3K36ac on an active (ACTIN7) and a silent (Ta2) gene. X-ChIP assays were performed using IgG, anti-H3 and anti-H3K36ac antibodies.

H3K36ac levels were high at the transcribed ACTIN7 gene but low at the Ta2 transposon. This observation is consistent with the immunolabeling results (Fig. 3.4) and confirms the preferred targeting of H3K36ac to transcribed genome regions.
Genome-wide distribution of H3K36ac on transcriptionally active genes

To localize H3K36ac throughout the genome, N-ChIP at a mono-nucleosome resolution was performed using chromatin from rosette leaves. The recovered DNA was subjected to Illumina DNA sequencing. A summary of the sequencing results is shown in Table 3.2. Nucleosome density was controlled for by normalizing H3K36ac signals to anti-histone ChIP signals. The overall distribution of H3K36ac along the five Arabidopsis chromosomes revealed high H3K36ac levels along all chromosome arms and low levels at the pericentromeric heterochromatin regions (Fig. 3.6). Furthermore, the H3K36ac signal (Fig. 3.6, red lines) correlated strongly with gene density (Fig. 3.6, green lines) consistent with an enrichment of H3K36ac at active genes. In contrast, H3K36ac showed a negative correlation with the density of transposable elements (Fig. 3.6, orange lines). These findings agree well with the results of the immunolocalisation of H3K36ac.

Table 3.2. Summary of Illumina sequencing of DNA isolated from the N-ChIP.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total reads</th>
<th>Mapped reads</th>
<th>Percent of total</th>
<th>Reads without excessive duplication</th>
<th>Percent of mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG_I</td>
<td>3,149,628</td>
<td>1,202,541</td>
<td>38%</td>
<td>1,137,543</td>
<td>95%</td>
</tr>
<tr>
<td>IgG_II</td>
<td>18,550,618</td>
<td>11,253,493</td>
<td>61%</td>
<td>8,177,197</td>
<td>73%</td>
</tr>
<tr>
<td>Input_I</td>
<td>24,624,908</td>
<td>23,520,553</td>
<td>96%</td>
<td>11,303,551</td>
<td>48%</td>
</tr>
<tr>
<td>Input_II</td>
<td>24,989,682</td>
<td>23,768,324</td>
<td>95%</td>
<td>6,817,167</td>
<td>29%</td>
</tr>
<tr>
<td>Histone_I</td>
<td>20,197,630</td>
<td>18,275,683</td>
<td>90%</td>
<td>14,266,456</td>
<td>78%</td>
</tr>
<tr>
<td>H3K36ac_I</td>
<td>22,027,778</td>
<td>19,962,283</td>
<td>91%</td>
<td>19,390,023</td>
<td>97%</td>
</tr>
<tr>
<td>H3K36ac_II</td>
<td>28,927,588</td>
<td>27,640,234</td>
<td>96%</td>
<td>26,677,265</td>
<td>97%</td>
</tr>
</tbody>
</table>
Figure 3.6. Genome wide distribution of H3K36ac on the Arabidopsis genome
Chromosomal distribution of H3K36ac (red) revealed by N-ChIP-seq was plotted together with gene density (green) and TE density (orange) in the Arabidopsis genome.

To establish the distribution of H3K36ac across genes, averaged H3K36ac profiles were plotted (Fig. 3.7 A). The results show a strong enrichment of H3K36ac within the first 500 bp downstream of the transcriptional start site (TSS). More specifically, the average H3K36ac peak was located between 35 bp and 516 bp downstream of the TSS (positions with 50% of maximal peak height). In contrast, very little H3K36ac was found in the promoter, further into the gene body or around the transcriptional termination site (TTS).
Next, we tested whether the width of the H3K36ac peak depended on gene length by plotting averaged H3K36ac profiles separately for bins of increasing gene length (Fig. 3.7 B). This analysis revealed that position, shape and width of the H3K36ac peak are independent of gene size, and that H3K36ac covered about 500 bp or three nucleosomes downstream of the TSS for very short as well as for very long genes. The invariant pattern of H3K36ac allowed calculating a gene-specific H3K36ac score as the mean H3K36ac signal within a window of 35 bp to 516 bp downstream of the TSS. The H3K36as score has a bimodal distribution that allows to separate 16297 genes with very low H3K36ac (score <=230) from 11131 genes with high H3K36ac (score >230). To avoid confounding effects of alternative transcripts, only genes with a single annotated transcript were included in this analysis (Fig. 3.7 C).
Next, we tested the relation between H3K36ac and transcription at a genomic scale. Generally, genes with higher mRNA levels had also higher H3K36ac scores (Fig. 3.8 A). Notably, however, this relation was not linear but strongly sigmoidal (red trend line in Fig. 3.8 B).

Figure 3. 8 Relationship between H3K36ac and transcriptional gene state (A) Relation between H3K36ac score and transcriptional state. Active genes with a high H3K36ac score have also a high transcript level, and inactive genes with very low H3K36ac score, have a low mRNA level. (B) H3K36ac plotted as function of transcript abundance with a running median trend line (red). Note the sigmoid relationship of the trend line. (C) Relationship between H3K36ac levels and the transcriptional state of diurnally expressed genes as well as with active and inactive not-diurnally regulated control genes. Expression data are from end of day when evening genes are highly expressed and non-evening genes are weakly expressed. Note that differences in H3K36ac are considerably larger between active and inactive control genes than between active and inactive diurnally regulated genes (see text for details). (D) Transcript abundance of diurnally expressed genes as well as active and inactive not-diurnally regulated control genes. Note that active and inactive control genes were selected for their similar expression levels to the diurnally expressed genes.
Thus, H3K36ac levels are only weakly proportional to transcription rate and instead are a qualitative indicator of transcriptional activity. This interpretation is consistent with the bimodal distribution of H3K36ac scores (Fig. 3.7 C). Then, we wondered whether H3K36ac rapidly follows changes in transcription. As a first test of this question we analyzed genes with strong diurnal expression changes. Diurnally regulated genes were taken from published data (Mizuno and Yamashino, 2008). Samples for ChIP-seq and transcriptome profiling were harvested at end of day, at zeitgeber time ZT=7, as described previously in (Shu et al, 2012); “Evening genes” were defined as diurnally expressed genes with an EOD expression signal >8; “non-evening genes” were defined as diurnally expressed genes with an EOD expression signal <6 (Fig. 3.8 C, D). For comparison, we defined two control sets of genes that are not diurnally regulated but have similar transcript levels as the evening or non-evening genes, respectively (Fig. 3.8 D). Consistent with the results described above, the H3K36ac scores differed greatly between the active and inactive control sets (Fig. 3.8 C). In contrast, H3K36ac levels differed only weakly between the evening and non-evening genes (Fig. 3.8 D). Specifically, H3K36ac was much lower for the active evening genes than for the active control genes and was much higher for the inactive non-evening genes than for the inactive control genes. Thus, while transcription can cycle rapidly with a diurnal period, H3K36ac seems to react much slower and “lack behind” transcription. These data suggest a model in which relatively high H3K36ac at inactive genes could constitute a memory for recent transcription while relatively low H3K36ac at active genes could constitute a memory for recent inactivity.

Because H3K36ac is associated with transcription, we tested its relation to other chromatin marks related to transcription, namely H3K4me3, H3K36me2, H3K36me3 and the histone variant H2A.Z. Averaged profiles around gene TSS were plotted separately for five bins divided by increasing H3K36ac scores. All tested chromatin marks showed a positive association with H3K36ac but the strongest association was found with H3K4me3 and H3K36me3 (Fig. 3.9).
Figure 3.9 H3K36ac positively correlates with active histone marks and H2A.Z at the TSS. Averaged relative enrichment of (A) H2A.Z, (B) H3K4me3, (C) H3K36me3 and (D) H3K36me2 for bins sorted by increasing H3K36ac scores.

H3K4me3, H3K36me3 and H2A.Z all were restricted to the gene 5’ end, while H3K36me2 was increasing towards the 3’ end. Next, we tested whether similar to H3K36ac the localization of the other chromatin marks was independent of gene length. Indeed, localization, shape and width of H3K4me3 and H2A.Z peaks were independent of gene length; the peaks remained always restricted to the gene’s 5’end (Fig. 3.10 A, B).
In contrast, the range of H3K36me2 and H3K36me3 strongly depended on gene length with H3K36me3 extending much further beyond the TSS for longer genes and H3K36me2 starting to increase much further beyond the TSS for longer genes (Fig. 3.10, C, D).

**Figure 3.10 Dependency of active histone marks on gene size.** Averaged relative enrichment of (A) H2A.Z, (B) H3K4me3, (C) H3K36me3 and (D) H3K36me2 was plotted around TSS for bins sorted by increasing gene size.

In order to better compare peak positions, averaged profiles for the analyzed marks were plotted together. This analysis revealed that H3K36ac together with H2A.Z are localized closest to the TSS and likely associated with transcription initiation and the transition to transcription elongation (H3K36ac peaks at...
195 bp and H2A.Z at 252 bp after the TSS). With transcription changing to elongation, H3K36ac and H2A.Z become less abundant and instead H3K4me3 (peak at 368 bp) and H3K36me3 (peak at 629 bp) transiently dominate. Eventually, H3K36me2, which is thought to associate with elongating Pol II, increases towards the 3’ end (Fig. 3.11).

**Figure 3.11 Averaged profiles of active marks around the TSS.** H3K36ac, Black; H2A.Z, Red; H3K4me3, Green; H3K36me3, Blue; H3K36me2, Light blue.

To confirm the *in silico* results, H3K36ac and H3K36me3 levels were investigated by ChIP qPCR for two test genes (Fig. 3.12). The obtained results confirmed that H3K36ac is highest at the 5’ end and decreased along the gene body. In contrast, no considerable change in the level of H3K36me3 was observed for the selected genes, suggesting that this modification may be less specifically associated with specific phase of transcription.
Figure 3. **H3K36ac is enriched at 5’ ends of active genes.** (A, B) The distribution of H3K36ac (light blue) and H3K36me3 (dark blue) across AT3G04720 and AT3G28060 based on ChIP-seq and ChIP-chip data, respectively. Gene structure and primer positions (region 1, 2, 3) at the bottom. (C, D) The relative enrichment of IgG, H3K36ac and H3K36me3 on the different regions of AT3G04720 and AT3G28060 using X-ChIP-qPCR with 10 days old Col seedlings. The enrichment relative to H3 signals is shown as mean ± S.E.M.

Taken together, H3K36ac is a highly enriched modification at the 5’ end of most transcriptionally active genes suggesting a potential role of this modification in the control of gene expression.
H3K36ac depends on GCN5 in Arabidopsis

Histone acetylation is established by HATs. In Arabidopsis 12 different HAT genes have been identified and grouped into four classes (Pandey et al, 2002), among which GCN5 (General Control Non derepressible protein 5) is the most abundant functional HAT in yeast, animals and plants. GCN5 is an evolutionary conserved protein, which in mammals is needed for a normal immune response and to repress tumor formation.

In plants, alteration of GCN5 has a pleotropic effects on vegetative and reproductive development (Servet et al, 2010, Vlachonasios et al, 2003). Because GCN5 is such an important HAT, we tested whether Arabidopsis GCN5 was required for the maintenance of H3K36ac. We used the T-DNA insertion gcn5-1 allele (Benhamed et al, 2008) for X-ChIP experiments with anti-H3 and anti-H3K36ac antibodies. The TSS region of selected genes was screened by qPCR. Indeed, the H3K36ac levels were up to five times lower in the gcn5-1 mutant than in wild type (Fig. 3.13 A, B). These results show that GCN5 is required for full H3K36 acetylation in vivo but that other HATs also contribute to acetylation of H3K36. The effect of GCN5 on H3K36ac could be direct or indirect. To test whether GCN5 can directly acetylate H3K36ac, in vitro HAT assays were performed using recombinant purified GST and GST-GCN5. The products of the acetylation reactions were run on gels, blotted and probed with anti-H3K36ac antibodies. A strong signal was obtained when using GCN5 enzyme while no signal was obtained with control GST or H3 protein alone (Fig. 3.13 C, D). Taken together these findings establish Arabidopsis GCN5 as a H3K36 acetyl transferase and strongly suggest that this enzyme generates most H3K36ac in vivo.
Figure 3. GCN5 is required for H3K36 acetylation. H3K36ac levels on region 1 (see Fig. 3.12) of AT3G28060 (A) and AT3G04720 (B). 10 days old seedlings from Col and gcn5-1 were used for X-ChIP with anti-H3 and anti-H3K36ac antibodies. The enrichment of H3K36ac relative to H3 is shown as mean ± S.E.M. (C-D) H3K36ac acetyltransferase activity of GCN5. Recombinant GST and GST-GCN5 were purified with Glutathione Sepharose 4B resin, blotted and probed with anti-GST antibodies; M, weight marker (C). HAT assays were performed using recombinant His-H3 as substrate. The reaction products were resolved on a 15% SDS-PAGE gel, transferred to PVDF membrane and probed with anti-H3K36ac antibodies. Anti-H3 was used as control.

HDA19 but not HDA6 controls H3K36 levels

In Arabidopsis, 18 different HDACs have been identified and grouped into three different classes. These proteins are evolutionary conserved and thought to act, at least some of them, redundantly (Pandey et al, 2002).
To identify potential HDACs responsible for H3K36 deacetylation, we tested the possible involvement of HDA6 and HDA19 - the two major Arabidopsis HDACs.

Figure 3. 14. HDA6 does not affect H3K36ac steady state levels. H3K36ac levels on region 1 (Fig. 3.12) of AT3G28060 and AT3G04720 were probed by X-ChIP using 10 days old seedlings from wild type, hda6 (B, D) and 35S::HDA6-FLAG (A, C). H3K36ac levels relative to H3 are shown as mean ± S.E.M.

HDA6 was reported to have activity against H3K9ac, H3K27ac, H3K14ac, H3K18ac, H3K23ac, H4K5ac, H4K8ac and H4K12ac (Hollender and Liu, 2008). To test a potential role of HDA6 in deacetylation of H3K36ac, X-ChIP assays were performed using wild-type, hda6 mutant (Aufsatz et al, 2002) and 35S::HDA6-FLAG overexpressing (Earley et al. 2006) plants. Two gene regions, where H3K36ac was previously found to be high, were tested.
No significant effects on H3K36ac levels were observed in HDA6 mutants or overexpressors (Fig. 3.14).
Taken together, these results suggest that HDA6 does not function to remove acetylation from H3K36.

Figure 3. 15. HDA19 affects H3K36ac levels

H3K36ac levels on region 1 (Fig. 3.12) of AT3G28060 and AT3G04720 were probed by X-ChIP using 10 days old seedlings from wild type and an HDA19 RNAi1 line (Zhou et al, 2005). H3K36ac levels relative to H3 are shown as mean ± S.E.M.

To test a potential role of the second main known Arabidopsis HDAC, HDA19, a knock-down line for HDA19 was used for X-ChIP assays, and the region where H3K36ac was found to be enriched before, was tested. A considerable increase in H3K36ac levels was observed in the HDA19 RNAi1 line (Fig. 3.15). This suggests that HDA19 can act as a deacetylase enzyme for H3K36ac.

H3K36ac is a conserved modification in plants

Histone proteins are among the most conserved eukaryotic proteins, and H3K36 is conserved in all known H3 proteins (Fig. 3.13). H3K36ac was found previously in animals and yeast (Morris et al, 2007). To test whether H3K36ac is restricted to Brassicaceae or present in other plants as well, bulk histones were extracted from various plant species including mono-and dicotyledonous angiosperms and a gymnosperm. The levels of acetylation at H3K36 were assessed by protein immunoblotting using anti-H3K36ac antibodies. A clear band was obtained with tobacco (dicotyledonous), rice (monocotyledon) as well as with spruce (gymnosperm).
The absence of a clear band in wheat extracts may be due to technical problems preventing the extraction of histone with H3K36ac or loss of acetylation during the extraction. The intensive band in rice may be due to the small genome size with high relative gene content. These results suggest that H3K36ac is a highly conserved modification in seed plants and is likely to have a conserved biological function.

Figure 3. H3K36ac is conserved in seed plants. (A) Amino acid alignment of the amino-terminal tails of H3 from mouse, yeast and several plants. (B) Acid-extracted histones prepared from tobacco (Nicotiana tabacum), wheat (Triticum aestivum), rice (Oryza sativa) and spruce (Picea abies) were resolved on 15% SDS-PAGE gels followed by immunoblotting using anti-H3K36ac antibodies. Anti-H3 was used as loading control.

DISCUSSION

With the aim of identifying additional histone marks in plant chromatin, a sensitive and specific mass spectrometry approach based on the CID/ETD (collision-induced fragmentation/Electron Transfer Dissociation) fragmentation technology (Syka et al, 2004) was used. This approach allows sequencing of highly charged longer peptides (>20 aa) and even intact proteins. This “top-down” analysis has earlier been successfully used to determine histone PTMs (Jufvas et al, 2011, Taverna et al, 2007, Xiong et al, 2010).

Using the meristematic tissue of cauliflower inflorescences, PTMs on H3 were identified. Many of the lysine acetylation and methylation events known in Arabidopsis were confirmed. Phosphorylated H3 peptides were not detected in this assay; this may be due to the lability and low abundance of H3 phosphorylation, which requires phospho-peptide enrichment steps during the histone extraction for detection (Zhang et al, 2007). Interestingly, modifications not described in plants so far were identified.
Among the new modifications, acetylation of K36 was chosen for further molecular and biochemical studies.

**H3K36ac is a conserved histone mark in seed plants**

H3K36ac was previously found in screens for additional histone modifications in yeast and Tetrahymena and found to be conserved in mammals (Zhang et al., 2012, Morris et al., 2007). Using high-specificity rabbit antibodies, we confirmed the mass spectrometric identification of H3K36ac in plants. In addition to cauliflower, H3K36ac was also detected in several other seed plants including the gymnosperm spruce and angiosperms such as Arabidopsis, tobacco, wheat and rice. The variable abundance of H3K36ac among the investigated species was possibly due to variable efficiency of histone extraction and the fast turnover of lysine acetylation. Conservation of H3K36ac among yeast, ciliates, animals and plants suggests a functional conservation.

**H3K36ac marks the 5’ end of transcribed genes in euchromatin**

H3K36ac was found mainly in euchromatin and excluded from chromocenters suggesting a connection to transcription. Genome-wide profiling of H3K36ac in Arabidopsis revealed a positive correlation with gene expression. We found H3K36ac specifically enriched within the first 500 bp downstream of the TSS. Even for very long genes, H3K36ac did not usually extend considerably outside of this range suggesting that this modification is associated with early phases of transcription but not with transcriptional elongation. This observation is in stark contrast to yeast, where H3K36ac was highest in promoters but mostly absent in the transcribed region (Morris et al., 2007). The peak of H3K36ac in Arabidopsis coincided well with a H2A.Z peak at gene 5’ ends. The different widths of the H2A.Z and H3K36ac peaks is possibly caused by the higher resolution of N-ChIP, which was used for H3K36ac profiling, compared to X-ChIP, which was used for H2A.Z profiling. N-ChIP has mononucleosome resolution while X-ChIP has 200-500 bp resolution, depending on the extent of sonication. The overlap of H2A.Z and H3K36ac suggests that nucleosomes with H3K36ac often contain H2A.Z.

It remains to be established whether incorporation or presence of H2A.Z favors acetylation of H3K36 and/or whether H3K36ac favors H2A.Z incorporation. Similar to H3K36ac and H2A.Z, H3K4me3 and H3K36me3 peak at the 5’ end of active genes. In the case of H3K4me3 the peak width is independent of gene length but the peak summit is shifted slightly towards the TTS. The H3K36me3 peak is shifted even further towards the 5’ end and its extension strongly depends on gene length. H3K36me2 finally increases after the H3K36ac peak and correlates with elongating Pol II.

Thus, H3K36ac, H2A.Z, H3K4me3, H3K36me3 and H3K36me2 form individual gradients along transcribed genes differentiating initiation, transition from initiation to elongation and transcriptional elongation.

We found that H3K36ac marks active genes but is not strongly proportional to transcription rates; rather, it marks transcribed genes in a qualitative manner with full acetylation reached even at moderate transcription rates. Interestingly, although H3K36ac is usually much higher at active than at inactive genes, this difference is much smaller for diurnally regulated genes. We interpret the relatively low H3K36ac levels at active diurnal genes as a sign that genes need to be continuously active for longer periods before full H3K36ac levels are established. Conversely, we interpret the relatively high H3K36ac levels at inactive diurnal genes as a sign that genes need to be continuously inactive for longer periods before H3K36ac is fully lost. This interpretation suggests that H3K36ac could form an epigenetic memory system for past transcriptional activity. More detailed experiments will be needed to test this hypothesis.
GCN5 and HDA19 maintain the H3K36ac homeostasis

The Arabidopsis genome encodes 12 HAT proteins, divided into four major groups GNAT, MYST, p300/CBP, and TAF1 (Pandey et al, 2002). GCN5 is a member of the GNAT (GCN5-related N-acetyltransferase) family. GCN5 has been highly conserved during evolution (Nagy and Tora, 2007, Servet et al, 2010). Among other substrates, GCN5 was reported to acetylate H3K14. Because the sequences surrounding K14 and K36 have some similarities (STGGK14AP versus STGGVK36KP) (Poux and Marmorstein, 2003), we tested a potential role of GCN5 to acetylate H3K36 both in vivo using ChIP assays and in vitro using HAT assays. Indeed, GCN5 can in vitro acetylate H3K36 and is in vivo required for full H3K36ac levels. Our results are consistent with previously reported acetylation of lysine 36 by GCN5 in yeast (Morris et al, 2007) and suggest a conserved H3K36 acetylation mechanism among eukaryotes.

Interestingly, however, the patterns established by the conserved mechanism differ between yeast and plants (see below).

The Arabidopsis genome encodes for 18 HDAC proteins, divided into three main subfamilies [RPD3, HD2, and SIR2 family] (Pandey et al, 2002). The Rpd3-like protein HDA6 functions in transgene silencing and the regulation of ribosomal RNA transcription. HDA6 is required to maintain DNA hypermethylation at the promoters of silenced rRNA genes. Disruption of HDA6 reduces levels of promoter DNA methylation and H3K9me2 but increases histone acetylation including H3K9ac and H3K14ac (Earley et al, 2006, To et al, 2011). The level of H3K36ac was not affected in HDA6 overexpressing and mutant lines, suggesting that HDA6 may not function in deacetylation of H3K36ac. The Arabidopsis RPD3-type HDAC HDA19 is another important HDAC in plant development (Benhamed et al, 2006). The level of H3K36ac increased in silenced HDA19 lines, suggesting that HDA19 may be required for maintenance of H3K36ac homeostasis on target genes in vivo.

H3K36ac/me2- an antagonistic pair with related function?

The genome-wide mapping of H3K36 trimethylation in yeast and mammals had revealed that this modification accumulates on the coding region of actively transcribed genes, peaking near the 3’ ends. In addition, H3K36me3 is a co-transcriptional modification, necessary for the activation of histone deacetylase complexes and plays a key role in the transcription elongation (Joshi and Struhl 2005, Pokholok et al. 2005, Venkatesh et al, 2012). Using higher resolution methods, H3K36me3 was found to be more enriched at exons than introns in Caenorhabditis elegans, mouse and humans (Kolasinska-Zwierz et al, 2009), supporting the hypothesis that transcription and splicing are coupled events (Sims and Reinberg, 2009). Reports from yeast suggested that H3K36me3 is recognized by the histone deacetylase complex (HDAC) RPD3S, through the chromodomain of Eaf3 (Joshi and Struhl, 2005). Subsequently, histone H3 and H4 get de-acetylated, thus preventing inappropriate nucleosome displacement at the 3’ end of genes and maintaining the dormancy of cryptic promoters in coding regions to suppress spurious intragenic transcription (Carrozza et al, 2005). These results indicated that histone deacetylation at coding regions is linked to H3K36 methylation, which in turn is linked to phosphorylation of the C-terminal domain of Pol II and the process of transcriptional elongation. (Joshi and Struhl, 2005).

Although conserved, modifications of H3K36 appear to have different functions in plants than in animals or yeast. In animals and yeast, H3K36me3 is the modification associated with transcriptional elongation, and H3K36ac in yeast is restricted to promoters. In plants, H3K36me2 is the modification associated with transcriptional elongation, and H3K36ac is restricted to the 5’ end of genes. It will be interesting to test whether these differences in H3K36 methylation and acetylation share common underlying mechanisms. It is tempting to speculate that acetylation and methylation at H3K36 form a molecular switch where one of the two dominates. Nevertheless, nucleosomes have two H3 molecules and concomitant acetylation and
methylation at H3K36 at the same nucleosome are possible. It remains to be tested whether such nucleosomes with asymmetrical K36 modifications indeed exist in plants. Although the function of H3K36ac is still unknown, it is tempting to speculate that in plants H3K36ac acts to exclude H3K36me2 from nucleosomes close to the TSS. Because Arabidopsis H3K36me2 is thought to share with yeast H3K36me3 the function to repress transcription initiation inside of genes, preventing H3K36me2 to spread to the TSS or even the promoter may be required for normal transcription of many genes.

ACKNOWLEDGMENT

We are very grateful to Keqiang Wua, to Craig Pikaard and to Dao-Xiu Zhou for kindly providing us seeds for HDA19-RNAi1, 35S::Flag-HDA6, gen5-1 and the pQE30-GCN5 construct, respectively. We thank the FGCZ for Illumina sequencing. This work was supported by grants from the Swiss National Science Foundation SNF and the Swedish Science Foundation VR as well as from the Swedish University for Agricultural Sciences.

REFERENCES


ROUDIER, F, AHMED, I, BERARD, C, SARAZIN, A, MARY-HUARD, T, CORTIJO, S, BOUYER, D, CAILLIEUX, E, DUVERNOIS-BERTHET, E, AL-SHIKHLEY, L, GIRAUT, L, DESPRES, B, DREVENSEK, S, BARNECHE, F, DEROZIER, S, BRUNAUD, V, AUBOURG, S,


Supplemental Table S1

List of gene-specific primers used for RT-qPCR assays

<table>
<thead>
<tr>
<th>Gene /Alias</th>
<th>Primer ID/ Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G10930, PP2A</td>
<td>DK053: TCTTAATTTGAATCGGTTACTAGG</td>
</tr>
<tr>
<td></td>
<td>DK054: CATAACAACAATTTCACAAGCA</td>
</tr>
<tr>
<td>AT5G09810, Actin7</td>
<td>LH1718: GGAACTGGAATGGTGAAGGCTG</td>
</tr>
<tr>
<td></td>
<td>LH1719: CGATTGGATACCTCAGAGTGAGGA</td>
</tr>
<tr>
<td>AT4G03740, Ta2</td>
<td>LH1252: ATACTCTCCACTTCCCCGTCTTCCTTTTA</td>
</tr>
<tr>
<td></td>
<td>LH1253: AAACGATGCGTTGGGATAGGTC</td>
</tr>
</tbody>
</table>

Supplemental Table S2

List of gene-specific primers used for ChIP-qPCR assays

<table>
<thead>
<tr>
<th>Gene/ primer position</th>
<th>Primer ID/ Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT3G04720_1&lt;sup&gt;st&lt;/sup&gt; region</td>
<td>LH1960: AGCAATAAGCACTCAGCTCTCA</td>
</tr>
<tr>
<td></td>
<td>LH1961: TCCGACCAACAACGGTCAGGCAA</td>
</tr>
<tr>
<td>AT3G04720_2&lt;sup&gt;nd&lt;/sup&gt; region</td>
<td>LH2004: TGGTCCACTATTCTCACACGTTAC</td>
</tr>
<tr>
<td></td>
<td>LH2005: TGGTCCACTATTCTCACACGTTAC</td>
</tr>
<tr>
<td>AT3G04720_3&lt;sup&gt;rd&lt;/sup&gt; region</td>
<td>LH2006: TACGAAACATATATCCCTCGAATAC</td>
</tr>
<tr>
<td></td>
<td>LH2007: GCCATTGATCCTGGATGATTT</td>
</tr>
<tr>
<td>AT3G28060_1&lt;sup&gt;st&lt;/sup&gt; region</td>
<td>LH1666: ATTCTGTCAAAGAGACGGCGGGAT</td>
</tr>
<tr>
<td></td>
<td>LH1667: ATAGAGTGCCACAGTGACGCCCA</td>
</tr>
<tr>
<td>AT3G28060_2&lt;sup&gt;nd&lt;/sup&gt; region</td>
<td>LH2008: GTGTGCGTTTCAAATCTCCTTGTG</td>
</tr>
<tr>
<td></td>
<td>LH2009: GCGACGATGCTATCACAACATAAT</td>
</tr>
<tr>
<td>AT3G28060_3&lt;sup&gt;rd&lt;/sup&gt; region</td>
<td>LH2010: TAGCGCTTTAGCCAAATCCA</td>
</tr>
<tr>
<td></td>
<td>LH2011: AAGGTCGTTGCCTCCATTAC</td>
</tr>
</tbody>
</table>
Supplemental Table 3

The 1-60 amino acid sequences of histone 3 from different organisms

> [Drosophila melanogaster]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE

> [Saccharomyces cerevisiae]
MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGTVALREIRRFQKSTE

> [Tetrahymena thermophila]
MARTKQTARKSTGAKAPRKQLASKAARKSAPATGGIKKPHRFPRPGTVALREIRKYQKSTD

> [Arabidopsis thaliana]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFPRPGTVALREIRKYQKSTE

> [Nicotiana tabacum]
RTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTE

> [Zea mays]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFPRPGTVALREIRKYQKSTE

> [Triticum spp]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFPRPGTVALREIRKYQKSTE

> [Oryza sativa]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFPRPGTVALREIRKYQKSTE

> [Picea abies]
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFPRPGTVALREIRKYQKSTE
H3K23me1 is an evolutionary conserved histone modification associated with gene silencing in Arabidopsis

Reference:
Manuscript in preparation

Authorship:
I performed most of the molecular and biochemical experiments and wrote the manuscript.
H3K23me1 is an evolutionary conserved histone modification associated with gene silencing in Arabidopsis

Mahrez, Walid1,2, Nanni, Paolo3, Gruissem, Wilhelm1 and Hennig, Lars1,2,4

Affiliations

1 Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, CH-8092, Zurich, Switzerland.
2 Swedish University of Agricultural Sciences, Department of Plant Biology and Forest Genetics and Linnean Center for Plant Biology, PO-Box 7080, SE-75007 Uppsala, Sweden.
3 Functional Genomics Center Zurich, University of Zurich/ETH Zurich, CH-8057 Zurich, Switzerland
4 Science for Life Laboratory, SE- 75105 Uppsala, Sweden

SUMMARY

Amino-terminal tails of histones are targets for diverse post-translational modifications, whose combinatorial action may constitute a code that will be read and interpreted by cellular proteins to define particular transcription states. Here, we describe monomethylation of H3 lysine 23 (H3K23me1) as a novel histone H3 modification in plants. H3K23me1 is an evolutionary conserved mark in diverse flowering plant species. It is found to be highly abundant on the heterochromatin enriched repetitive DNA sequences and retroelements. Using ChIP, H3K23me1 was found enriched on silent Polycomb target genes and on constitutively heterochromatin located genes. The histone methyltransferase KRYPTONITE was required for H3K23 monomethylation. H3K23me1 on transposable elements did also depend on normal DNA methylation. Together, this work establishes H3L23me1 as a novel component of the plant histone code.
INTRODUCTION

Histones are small, highly basic proteins. Two pairs of H3, H4, H2A and H2B constitute together an octamer, around which about 147 bp of DNA is wrapped to constitute the basic unit of chromatin – the nucleosome. The amino-terminus of histones can carry several covalent post-translational modifications (PTMs) such as acetylation, phosphorylation, methylation, ubiquitination and ADP-ribosylation (Bannister and Kouzarides, 2011). These dynamic and often evolutionary conserved modifications act in combination to establish distinct chromatin states, thus constituting a histone code (Jenuwein and Allis, 2001). Histone PTMs can be read and interpreted by specific proteins, which subsequently influence numerous biological processes including transcription, replication, chromosome maintenance and cell division.

Unlike histone lysine acetylation which can affect the DNA-protein interaction, histone lysine methylation is an epigenetic modification that does not alter the charge of the modified lysines on histone tails. Lysines can be mono- (me1), di- (me2) or trimethylated (me3) on their ε-amine group. Using mass spectrometry, methylation of the histone H3 lysines K4, K9, K27, K36, K79 and the H4 lysine K20 was found (Liu et al, 2010a).

Depending on the target lysine and the degree of methylation, methylation can be associated with active or inactive genes and can have different cytological localization (Liu et al, 2010a). Although these patterns are often conserved, there are many differences between different groups of eukaryotes. In mice, for example, H3K9me3 and H3K20me3 are most abundant in condensed heterochromatin and associated with gene silencing whereas H3K9me1 and H3K9me2 are enriched in euchromatin (Loidl, 2004). In contrast, in Arabidopsis H3K9me3 and H4K20me3 are located in euchromatin and associate with gene expression (Charron et al, 2009) while H3K9me1 and H3K9me2 are localized in heterochromatin and associated with gene silencing (Fransz et al, 2006). Histone lysine methylation has important roles in many biological processes including control of transcriptional, cell-cycle regulation, DNA damage and stress responses, heterochromatin formation and X-chromosome inactivation (Greer and Shi, 2012, Martin and Zhang, 2005).

The homeostasis of the methylation on histone lysines is maintained by histone lysine methyl transferases (HKMT), which catalyze the transfer of methyl groups from S-adenosylmethionine to lysines’ ε-aminogroup, and by histone demethylases, which remove the methylation mark (Liu et al, 2010a, Black et al, 2012).

Histone lysine methylation is catalyzed by at least two HKMT families: SET domain-containing proteins (Rea et al, 2000) and DOT1-like proteins (Feng et al, 2002). H3K9me2 in Arabidopsis, for instance, is generated by the SET domain protein KRYPTONITE (KYP) (Jackson et al, 2004). Histone lysine methylation can be removed by two families of histone demethylases: FAD-dependent monoamine oxidases, which can demethylate only mono- and dimethylated lysines, and jumonji C (JmjC)-domain-containing, iron-dependent dioxygenases, which can demethylate even trimethylated lysines (Chen et al, 2011, Spedaletti et al, 2008). Enzymes of these groups are highly conserved in yeast, plants and animals. The disruption of these enzymes often leads to developmental defects or diseases such as cancer or cognitive disorders (Greer et al, 2012).

Once established, methylated lysines constitute landmarks for recognition by the aromatic cages of proteins with methyl-binding domains (Taverna et al, 2007) such as PHD fingers, WD40 repeats, CW domains (Hoppmann et al, 2011), PWWP domains, ankyrin repeats (Collins et al, 2008), chromodomains, double chromodomains, chromobarrels, Tudor domains, double or tandem Tudor domains and MBT
repeats. These protein readers often recruit additional protein complexes to perform functions such as gene silencing (Greer and Shi, 2012).

Genome-wide profiling using ChIP-chip or ChIP-seq in Arabidopsis (Roudier et al, 2011), Drosophila (Kharchenko et al, 2011) and in mammals (Ernst and Kellis, 2011), have led to the definition of “chromatin states”, distinguishable by their histone methylation patterns. For example, four main chromatin states were defined in Arabidopsis, in which three states were defined by their enrichment of a group of methylation marks. Active genes, enriched in H3K4me3 and H3K36me3, constitute CS1. CS2 and CS3 are defined by distinct types of repressive chromatin PTMs; the H3K27me3 associated to Polycomb group proteins, and the classical heterochromatin defining PTMs H3K9me2 and H4K20me1, respectively (Roudier et al, 2011).

The addition of a methyl group to cytosine bases of DNA to form 5-methylcytosine, referred to as DNA methylation, is another layer of epigenetic control that can stabilize repressed chromatin domains. DNA methylation occurs in both prokaryotes and eukaryotes. Although absent in budding and fission yeast as well as in Caenorhabditis elegans, this mark is prominent in fungi, plants and vertebrates (He et al, 2011). In plants, DNA methylation occurs in the contexts of CG, CHG and CHH (H = A, C, or T) and is highly abundant on centromeric and pericentromeric regions as well as other repetitive elements.

DNA methylation in the CG sequence context also occurs in the transcribed region of nearly one-third of expressed Arabidopsis genes (Law and Jacobsen, 2010). In plants three different DNA methyl transferases are well characterized: METHYLTRANSFERASE1 (MET1), CHROMOMETHYLASE3 (CMT3), DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Law and Jacobsen, 2010). Several studies have revealed the complex relationship between DNA methylation and histone PTMs. In Arabidopsis, loss of MET1 causes a significant reduction of H3K9me2 at heterochromatic loci. Conversely, the reduction of the H3K9me2 levels in kyp mutants cause hypo-methylation on many transposable elements (Zhou, 2009). In vitro assays have revealed the binding of KYP to GC methylated DNA by MET1 protein via its SRA (SET and RING associated) domain (Johnson et al, 2007). Moreover, once H3K9me2 is established by KYP, it can act as a binding site for CMT3, which subsequently methylates cytosines in CHG contexts (Law and Jacobsen, 2010). Furthermore, genome wide profiling has revealed a strong correlation between DNA and H3K9 methylation (Bernatavichute et al, 2008). In addition, it was found that DNA methylation coexists not only with H3K9me2 but also with H4K20me1, H3K27me1 and H3K27me2 (Roudier et al, 2011). This suggests a cooperative role of DNA and histone lysine methylation in transcriptional gene regulation.

Despite the detailed knowledge about many histone PTMs, new modifications are continuously discovered demonstrating the limitations of current knowledge (Tan et al, 2011, Fujiki et al, 2011, Chen et al, 2007, Bergmuller et al, 2007). Compared to yeast and animals, the histone PTMs in plants are even less well-defined. For a comprehensive understanding of chromatin function and for future modeling approaches, comprehensive lists of histone PTMs need to be established at least for model organisms. In a previous study we had used high-sensitivity mass spectrometry to probe for additional PTMs on the plant H3 amino-terminal tail. We had identified H3K36ac, which has not yet been characterized in plants. Here, I describe H3K23me1 as a new plant histone PTM. H3K23me1 is evolutionary conserved among flowering plants, highly enriched in the heterochromatin but also at other silent genes. KYP is required for H3K23me1 accumulation. In addition, H3K23me1 was found to depend on DNA methylation.
MATERIALS AND METHODS

Plant material and growth conditions

The Arabidopsis thaliana (L.) Heynh wild-type accession Columbia (Col), tobacco (Nicotiana tabacum), wheat (Triticum aestivum), rice (Oryza sativa) and spruce (Picea abies) were used. The kyp-lh1 allele (SALK_105816) was obtained from the Nottingham Arabidopsis Seed Stock Centre. Seeds were sown on 0.5× basal salts Murashige and Skoog (MS) medium (Duchefa, Haarlem, The Netherlands), stratified at 4°C for one day and then transferred to growth chambers for germination at 20°C under long day (LD, 16 h light) photoperiods for ten days. The seedlings were transferred into pots containing soil and left to grow in growth chambers under LD conditions. Wheat and rice were grown in a green house at 22°C under long day (LD, 16 h light) photoperiods. For Rice, special watering conditions were applied. Cauliflower was obtained from a local supermarket. Norway spruce (Picea abies) needles were collected from a ~15 years old wild-grown tree in a forest close to Uppsala (Sweden) located at 59.81° N, 17.66° E. For treatments with the DNA methylation inhibitor Zebularine (1-(β-d-ribofuranosyl)-1,2-dihydropyrimidine-2-one) (AH Diagnostics, Solna, Sweden), sterilized Arabidopsis seeds were germinated on MS medium containing 80 μM Zebularine. After 10 days, seedlings were collected for RNA extraction and ChIP assays.

RNA isolation and RT-qPCR

RNA extraction and reverse transcription were performed as described previously (Alexandre et al., 2009) with minor modifications. The DNA-free RNA was reverse-transcribed using a RevertAid First Strand cDNA Synthesis Kit (Fermentas, Helsingborg, Sweden) according to manufacturer’s recommendations. Aliquots of the generated cDNA were used as template for quantitative real-time PCR (qPCR) with gene-specific primers (Supplemental Tab. S1). qPCR was performed using SYBR green (Fermentas) on an IQ5 multicolor Real time PCR cycler (BIO-RAD, PA, USA). qPCR reactions were performed in triplicate, gene expression levels were normalized to a PP2A control gene, and results were analyzed as described (Simon, 2003).

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were performed as described (Exner et al., 2009). Approximately, 1.5 g of ten-days-old cross-linked whole seedlings were used. Antibodies used for ChIP were anti-H3 antibody (#07-690, Upstate/Millipore); anti-H3K9me2 (#pAb-060-050, Diagenode, Belgium); anti-H3K23me1 (#39387, ActiveMotif) and non-immune IgG #15006 (Sigma-Aldrich). qPCR was performed using SYBR green (Fermentas) and gene-specific primers (Supplemental Tab. S2) on an IQ5 multicolor Real time PCR cycler (BIO-RAD). qPCR was performed with three technical replicates and results were normalized to input, and presented as the enrichment level of the given modification over the level of histone H3.

Immunostaining

Immunostaining was performed according to (Jasencakova et al., 2000(Jasencakova et al., 2000) with some modifications. Paraformaldehyde prefixed Arabidopsis root tips were washed in MTSB buffer [50 mM PIPES, 5 mM MgSO₄, and 5 mM EGTA, pH 7.9], and digested for 10 min at 37°C with a PCP enzyme mixture [2.5% pectinase, 2.5% cellulase Onozuka R-10, and 2.5% Pectolyase Y-23 (w/v) dissolved in MTSB].
After washing with MTSB, root tips were squashed on a glass slide using a cover slip, and immediately frozen in liquid nitrogen. The cover slip was gently removed, and the nuclei were covered with blocking solution (MTSB containing 3% BSA) for 1 hour at 4°C. Next, the samples were incubated for 1 h at 4°C with anti-H3K23me1 #39387 (ActiveMotif) primary antibody in MTSB (containing 1% BSA and 0.1% Tween 20) and kept in a humid chamber overnight. After 3 times washing with MTSB for 5 min each, the slides were covered for 2 h at 4°C with Rhodamine-conjugated anti-rabbit IgG antibody (Thermo Scientific) diluted in MTSB containing 1% BSA and 0.1% Tween 20. The slides were gently washed with MTSB, and the DNA was counterstained with 1 μg/ml of DAPI in mounting medium (Vectashield; Vector Labs). Fluorescence signal detection and documentation was performed with a Leica DMI 4000 microscope. Images were merged and processed using Photoshop and ImageJ.

**Histone extraction form plants, fractionation by RP-HPLC and PTM identification by Tandem Mass Spectrometry**

Histone extraction form plants, fractionation by RP-HPLC and PTM identification by Tandem Mass Spectrometry were described previously (Chapter 3.2).

**Dot blotting**

For Dot blotting, 1µg of acid extracted histones, dissolved in water was blotted onto a Nitrocellulose membrane Hybond ECL™ (Amersham Life Science, Stockholm, Sweden), and left to air dry for approximately 1h. The membrane was then blocked with 5% nonfat milk in TBST, and probed with anti-H3K23me1 (#39387, ActiveMotif) antibodies.

**Sequence alignment of H3 amino-terminal tails**

Amino acid sequences of H3 from representative eukaryotes were selected using PSI-BLAST searches. Sequences were aligned using the ClustalW multiple sequence alignment program implemented in MEGA5 (Tamura et al, 2011). Visualization of the sequence alignment was performed with GenDoc; the amino acid sequences in the FASTA format are available in Supplemental Tab. S3.
RESULTS
H3K23me1 is a new modification in plants

We had performed a survey of histone H3 PTMs in plants (chapter 3.2). The manual validation of the chromatograms led to the identification of H3K23me1, which has not been described in plants before. The MS/MS spectrum of the ArgC digested histone H3 in Figure 3.1 shows the C and Z fragmentation ions for the parent peptide with an observed mass of m/z 999.62.

<table>
<thead>
<tr>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
<th>c7</th>
<th>c8</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
<td>274</td>
<td>387</td>
<td>458</td>
<td>559</td>
<td>701</td>
<td>772</td>
<td>843</td>
</tr>
</tbody>
</table>

K18 Q L A T K23me A A R26

856 (K18)  728 (Q)  615 (L)  544 (A)  443 (T)  301 (K23me)  159 (A)  230 (R26)

K18QLATK23meAAR26

Figure 3.1. Identification of H3K23me1 from Brassica oleracea MS/MS fragmentation spectrum of the [M+2H]²⁺ parent ion at m/z 999.62. This peptide was identified as the H3 peptide (histone H3, amino acids 18-26) derived from ArgC digested RP-HPLC-purified cauliflower H3. The y-axis represents the relative abundance of the parent ion; x-axis represents the mass/charge of the parent ions. Above the spectrum is the peptide sequence in which predicted c-type ions, which contain the N terminus of the peptide, are immediately above the sequence. Predicted z-type ions, which contain the C terminus, are immediately below the sequence; ions observed in the spectrum, which represent masses associated with the fragmented peptides from the MS/MS analyses, are underlined.

The accurate mass of the parent ion recorded was found to be consistent with a monomethylation on this peptide. Importantly, we were also able to detect H3K23 acetylation on other H3 peptides, confirming that this residue can be methylated or acetylated.

To verify the mass spectrometry results, histone extracts from Arabidopsis inflorescences and cauliflower seedlings and inflorescences were subjected to dot blotting using a polyclonal anti-H3K23me1 antibody.
This antibody was previously tested for specificity using peptide arrays as described (Liu et al, 2010b). Positive signals were obtained from all analyzed tissues and organisms (Fig. 3.2 A).

**Figure 3.2.** H3K23me1 is present in both Arabidopsis (inflorescences) and Cauliflower (seedlings, inflorescences). A) Dot-blot analysis of 1 µg of acid extracted histones from Arabidopsis thaliana (At) inflorescences and Brassica oleracea (Bo) inflorescences and seedlings using anti-H3K23me1 antibodies. B) Acid-extracted histones prepared from Arabidopsis thaliana (At) inflorescences were resolved on a 15% SDS-PAGE gel followed by protein immunoblotting using rabbit anti H3K23me1 and anti H3 antibodies.

To exclude any non-specific cross-reactivity of the antibody with other plant proteins, an immunoblot was performed with a histone extract from Arabidopsis inflorescences. A single band at ~17 kDa corresponding to the H3 size was obtained, confirming the dot blot results (Fig. 3.2B). These results confirm the existence of H3K23me1 in Arabidopsis and suggest a conservation of this modification among Brassicasae.
H3K23me1 is a conserved modification in plants

Histones are among the most strongly conserved eukaryotic proteins. H3K23me1 was found previously in mammalian cells (Liu et al, 2010b) and in the ciliate Tetrahymena (Zhang et al, 2012). To investigate the degree of conservation of lysine 23 among other eukaryotes; a sequence alignment of the first 60 amino acids of H3 from Arabidopsis and several other representative eukaryotes was performed (Fig. 3.3 A). This analysis revealed a strong conservation of lysine 23.

Figure 3.3 H3K23me1 is a conserved modification in all plant kingdoms

(A) Amino acids alignment of the amino-terminal tail of H3 from mouse, yeast and several plants. (B) Acid-extracted histones prepared from tobacco (Nicotiana tabacum), wheat (Triticum aestivum), rice (Oryza sativa) and spruce (Picea abies) were resolved on 15% SDS-PAGE gels followed by immunoblotting using anti-H3K36ac antibodies. Anti-H3 was used as loading control.

To test whether H3K23me1 exists in other plant species than Arabidopsis and cauliflower, bulk histone extracts were obtained from different plant species including mono- and dicotyledonous angiosperms and gymnosperms. The level of methylation at H3K23 was assessed by immunoblotting using anti-H3K23me1 antibodies. A clear band was obtained with all analyzed species (Fig. 3.3 B). These results show that H3K23me1 is a highly conserved modification among seed plants and suggest important biological functional roles for this modification.
H3K23me1 is enriched in heterochromatin

To investigate the localization of H3K23me1 in the nucleus, paraformaldehyde-fixed Arabidopsis root-tip nuclei were immunostained using anti-H3K23me1 antibodies. The H3K23 methylation fluorescence signal was present throughout euchromatic regions. However, H3K23me1 was strongly enriched in the DAPI-dense chromocenters, which contain centromeric and pericentromeric heterochromatic sequences (Fig. 3.4). This observation was similar to the staining pattern observed for H3K9me2 (Jackson et al., 2004) and suggests that H3K23me1 is mainly targeted to condensed and transcriptionally silent regions.

Figure 3.4. Localization of H3K23me1 in root tip nuclei.
Root tip interphase nuclei of Arabidopsis plants were stained with DAPI and analyzed for immunolocalization with anti-H3K23me1 antibody. A) DAPI, B) H3K3me1, C) overlay. A section through the nucleus was analyzed using ImageJ software, D) Quantitative line profiles of DAPI (blue), and H3K23me1 (red) fluorescence intensities along the white line shown in (C). Scale bar: 5 µm
**H3K23me1 is enriched on Polycomb group protein target genes and transposable elements**

To investigate the localization of H3K36ac in chromatin, ChIP assays were performed. Chromatin preparations from 11-days-old cross-linked Arabidopsis seedlings were subjected to immunoprecipitation using anti-H3K23me1, anti-H3K9me2 and anti-H3 antibodies. The purified, eluted DNA was evaluated by qPCR using primers targeting a constitutively active gene (ACTIN7), silent Polycomb group protein (PcG) targets (AG, SEP3), a silent non-PcG target (AT2G24370) and heterochromatin markers (Ta1, AtCOPIA4) (Fig. 3.5).

![Figure 3.5](image)

**Figure 3.5. H3K23me1 is enriched on silent genes** The relative enrichment of IgG, H3K9me2 and H3K23me1 on Actin7, AG, SEP3, AT2G42600, Ta (At4g03760), and AtCOPIA4 (AT4G16870) using ChIP-qPCR. The enrichment of the different histone marks relative to H3 signal is shown as mean ± S.E.M.

H3K23me1 enrichment relative to H3 levels was assessed. The active gene (ACTIN7) had only background signals of H3K9me2 and low levels of H3K23me1. The silent PcG targets did not carry any H3K9me2 but had some H3K23me1; similarly H3K23me1 was also increased at a silent, non-PcG target gene (AT2G42600). The typical heterochromatic transposons Ta1 and AtCOPIA4 carried much H3K9me2 and H3K23me1. H3K23me1 levels were considerably higher at the heterochromatic transposons than at the silent euchromatic genes. Together, H3K23me1 was found on diverse silent genes with the strongest enrichment at heterochromatic sequences where it co-localized with H3K9me2.

**H3K23me1 depends upon KYP in Arabidopsis**

Histone methylation is established by HKMTs. KYP is an H3K9me2 methyltransferase in Arabidopsis (Jackson et al., 2004). Because of the co-localization of H3K23me1 with H3K9me2 on transposable elements and the enrichment of H3K23me1 in heterochromatic chromocenters, we tested a potential role of KYP in the monomethylation of H3K23. To this aim, 10-days-old kyp mutant seedlings were used for ChIP using anti-H3, anti-H3K9me2 and anti-H3K23me1 antibodies. In kyp mutants, H3K9me2 at transposons was almost completely lost (Fig. 3.6). Interestingly, H3K23me1 was also reduced in kyp although only up to half of wild-type levels.

115
These results show that KYP is required for normal level of H3K23 monomethylation \textit{in vivo}. It remains to be tested whether KYP can directly methylate H3K23 or whether it affects H3K23 methylation indirectly.

![Figure 3.6. KYP is required for H3K23 methylation](image)

The relative enrichment of IgG, H3K9me2 and H3K23me1 on the transposable element genes \textit{Ta1} (\textit{At}4g03760) and \textit{Ta2} (\textit{AT}1G35960) using ChIP-qPCR in Col and \textit{kyp-ih1}. The enrichment of the different histone marks relative to H3 signal is shown as mean ± S.E.M.

**DNA methylation and H3K23me1 are interrelated silencing marks**

DNA methylation at cytosine bases is an important epigenetic modification establishing heterochromatic gene silencing. The functional relationship between DNA methylation and H3K9me2 is well-characterized in silencing of repetitive sequences in centromeric and pericentromeric regions of the Arabidopsis genomes.

In an attempt to test a potential functional interdependence between DNA methylation and H3K23me1, 10-days-old seedlings germinated on a medium supplemented with 80 μM of the stable DNA methyltransferase inhibitor Zebularine, were used to assess the expression of selected transposable elements (TE) as well as the enrichment of H3K9me2 and H3K23me1. As previously reported (Baubec et al, 2009); TE expression in non-treated control plants was very low (Fig. 3.7 A, C, E).

In plants treated with Zebularine, however, expression was considerably increased, demonstrating the release of silencing induced by the DNA hypomethylating drug. To test the effect of the DNA hypomethylation on the level of H3K23me1, a ChIP assay was performed on the treated and non-treated samples.

Zebularine lead to decreased H3K9me2 levels at one of the three tested TEs but had no effect on the other two (Fig. 3.7 B, F, D). In contrast, H3K23me1 was consistently reduced to about half at all three TEs. Taken together, these results suggest a functional relationship between the DNA methylation and H3K23me1 at the selected analyzed genes, and thus a possible cooperative effect in the transcriptional genes silencing. Because changes in H3K23me1 were found when H3K9me2 was not altered, these results also argue against the model that H3K9me2 is sufficient to recruit H3K23me1.

116
Figure 3. 7. A relationship between DNA methylation and H3K23 methylation

(A, C, E) Expression level of \textit{AtMu1}, \textit{AtSN1} and \textit{Ta1} (\textit{At4g03760}) in Col and Col treated with 80 μM Zebularine. RNA was extracted from 10 day-old seedlings grown at LD conditions; cDNA was synthesized with oligo dT primers. Relative expression to \textit{PP2a} is shown as mean ± S.E.M. (n = 3). (B, D, F) H3K9me2 and H3K23me1 relative enrichment on \textit{AtMu1}, \textit{AtSN1} and \textit{Ta1} (\textit{At4g03760}) using ChIP-qPCR. Ten days old seedlings from Col and Col treated with 80 μM Zebularine were used for X-ChIP. The enrichment of the different histone marks relative to H3 signal is shown as mean ± S.E.M.
DISCUSSION

In the present study we establish H3K23me1 as a histone H3 PTM in seed plants. H3K23me1 was detected by mass spectrometry in cauliflower and by immunoblotting in cauliflower, Arabidopsis, tobacco, wheat, rice and spruce. Earlier reports described H3K23me1 in human cell lines and in Tetrahymena (Liu et al, 2010b, Zhang et al, 2012, Leroy et al, 2012).

It remains to be tested whether H3K23me1 exists in insects, yeast and lower plants as well. In Arabidopsis, H3K23me1 was enriched in heterochromatic chromocenters and present on TE sequences. This is consistent with the observation that H3K23me1 colocalizes with the heterochromatin marker HP1β in mammals (Liu et al, 2010b) and is bound by mammalian HP1α and HP1β in vivo and in vitro (Liu et al, 2010b; Leroy et al, 2012).

In Arabidopsis, heterochromatin is marked by H3K9me2 (Jackson et al, 2004), which is produced by the HKMT KYP (Jackson et al, 2004). We tested the relation between KYP, H3K9me2 and H3K23me1 using kyp mutants. In contrast to H2K9me2, which was almost entirely lost, H3K23me1 was reduced only to ~50% in kyp. Thus, KYP is needed for H3K23me1. Furthermore, the partial reduction of H3K23me1 in kyp, suggest the potential redundant action of KYP with other HKMTs to produce H3K23me1. SUVH5 and SUVH6, two close KYP/SUVH4 homologs, are likely candidates (Ebbs and Bender, 2006). Similarly, two SET-domain containing proteins ATXR5 and ATXR6, were found to act together to deposit the H3K27me1 over transposable elements sequences in Arabidopsis (Jacob et al, 2009). Alternatively, it is possible that presence of H3K9me2 favors the recruitment of an unknown HKMT that adds H3K23me1 to the local chromatin.

In addition to H3K9me2, DNA methylation is another marker of heterochromatin in Arabidopsis (He et al, 2011). To study DNA methylation, cytidine analogs, which specifically inhibit the catalytic activity of DNA methyl transferases, are often used. 5-azacytidine (5-azaC) is one of the widely used inhibitors of DNA methylation. However, this drug is relatively instable and often toxic. Therefore, the more stable and less toxic inhibitor Zebularine was recently introduced as an efficient mean to block DNA methylation (Zhou et al, 2002, Cheng et al, 2003). This drug was successfully used in fungi, ciliates as well as in plants (Cheng et al, 2003, Malik et al, 2012, Baubec et al, 2009). We used Zebularine to test a potential relationship between H3K23me1 and DNA methylation. In contrast to H3K9me2, which was reduced on only some genes, H3K23me1 dropped at all analyzed loci. This suggests that DNA methylation can recruit H3K23me1.

It remains to be tested whether H3K23me1 in turn can recruit DNA-methylation as previously shown for H3K9me2 (Du et al, 2012). Together, H3K23me1 is a component of heterochromatin in Arabidopsis and depends, at least partially, on undisturbed H3K9me2 and DNA methylation.

In addition to the high H3K23me1 levels in heterochromatin, we found H3K23me1 at intermediate levels also on other silent genes such as inactive PcG target genes. It remains to be tested whether H3K23me1 in mammals is also present at inactive PcG targets or whether this is a plant-specific trait. The silencing of PcG target genes is mediated by the deposition of H3K27me3. This repressive mark, which controls the expression of ~17% of coding genes in Arabidopsis, defines by itself a distinct chromatin state (Turck et al, 2007, Roudier et al, 2011, Zhang et al, 2007). AG and SEP3 are floral homeotic genes, which control flower development and are expressed only in flowers. Their ectopic expression in non-flower tissue is prevented by PcG protein action. Interestingly the increased level of H3K23me1 on AG and SEP3 suggests a potential cooperative role of this mark with H3K27me3 in maintaining stable silencing of target genes.
Notably, the plant PcG system involves LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), a homolog of the mammalian HP1 that directly binds AG and SEP3 (Exner et al, 2009). Because HP1 binds H3K23me1 (Liu et al, 2010b), it is tempting to speculate that in plants H3K23me1 contributes to recruit LHP1 to PcG targets. Interestingly, mutants in Tetrahymena EZL2, a homolog of the PcG H3K27me3 HKMT E(Z), showed a compensatory increase in H3K23me1 (Zhang et al, 2012). Thus, the relation of H3K23me1 to PcG protein function remains to be analyzed in more detail.

Generally, H3K23me1 was associated with transcriptional repression. Future work has to establish whether H3K23me1 directly contributes to gene silencing and which chromatin readers bind to H3K23me1 in plants to convey regulatory effects.

ACKNOWLEDGMENT

We are very grateful to Drs. Sun Chuanxin, Joel Sohlberg, and Mohammad Sameri, for kindly providing us seeds for rice and leaf material for spruce and wheat, respectively. This work was supported by grants from the Swiss National Science Foundation SNF and the Swedish Science Foundation VR as well as from the Swedish University for Agricultural Sciences.

REFERENCES


KHARCHENKO, P. V, ALEKSEYENKO, A. A, SCHWARTZ, Y. B, MINODA, A, RIDDL, N. C, ERNST, J, SABO, P. J, LARSCHAN, E, GORCHAKOV, A. A, GU, T. T, LINDER-BASSO, D,


### Supplemental Table S1

List of gene-specific primers used for RT-qPCR assays

<table>
<thead>
<tr>
<th>Gene /Alias</th>
<th>Primer ID / Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G10930, PP2A</td>
<td>DK053: TCTTAATTTGAATCGGTTACTAGG DK054: CATACTAAACAATTTGTTGCAAGCA</td>
</tr>
<tr>
<td>AT3RE63860, AtSN1</td>
<td>HJ12: ACCAACGTGGTGGGGCCAGTGG HJ13: GAAATCTCTGAAATCGTCAACAGGC</td>
</tr>
<tr>
<td>AT4E23190, AtMUI</td>
<td>MD050: TTGTGAGCATTCTTATGGTGCA MD051: TGTCGAGAAGAACCTGGAATATG</td>
</tr>
<tr>
<td>AT4G16870, ATCOPIA4</td>
<td>LH2030: CTTGTTTTGCTTTCCCCCTGTGT LH2031: TAGCGAAGAAGCTCCTGCA</td>
</tr>
<tr>
<td>AT4G03760, TRANSPONSON</td>
<td>MD054: CTTGTATGAGGGGTTCTTTATGG MD055: AGAACCAGCCTGAGGTATA</td>
</tr>
<tr>
<td>AT1G35960, TRANSPOSON</td>
<td>PW236: AACAACCTAGAGAAGGTTCCGA PW237: TCCACTTCCATTTGGAACCA</td>
</tr>
</tbody>
</table>

### Supplemental Table S2

List of gene-specific primers used for ChIP-qPCR assay

<table>
<thead>
<tr>
<th>Gene /Alias</th>
<th>Primer ID / Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT5G09810, ACTIN 7</td>
<td>LH1255: AGCGAACCAGGATCTAGAGACTCACCTTG LH1254: CGTTCGCTTTCTTCTATGTGTTAGCT</td>
</tr>
<tr>
<td>AT1G24260, SEP3</td>
<td>LH0666 : ATGGATCTTTGCTCTTCTATCAACAG LH0667: AGAGAGAGATTGAGATATCTTCTTGG</td>
</tr>
<tr>
<td>AT2G24370, AG</td>
<td>LH0658 : CTAATCAAAATTTGCGCTAAAACG LH0659 : TCCATAGCCTCGATGGTACAG</td>
</tr>
<tr>
<td>AT2G24370, --</td>
<td>LH2028: TGGGGATGTTTAGCTACGAGTTC LH2029: CATGATGTTTGTGCTTAGTCA</td>
</tr>
<tr>
<td>AT3RE63860, AtSN1</td>
<td>HJ12: ACCAACGTGGTGGGGCCAGTGG</td>
</tr>
</tbody>
</table>
Supplemental Table S3

The 1-60 amino acid sequences of histone 3 from different organisms

>Saccharomyces cerevisiae
MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGTVALREIRRFQKSTE

>Arabidopsis thaliana
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE

>Nicotiana tabacum
RTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTE

>Triticum spp
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE

>Oryza sativa
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE

>Picea abies
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE
4. Concluding discussion

Flower formation and subsequently seed production are important steps during flowering plant development, as this maintains the species through generation and maintains a continuous supply of seed crops to feed a growing world population. Flowering time is a very important agronomical trait, which has a large impact on crop yield. Plant breeders have a strong interest in understanding the control of flowering and the mechanisms of plant adaptation to changing environments. This interest has intensified attempts to unravel the molecular mechanisms underlying the control of flowering time in plants. Despite the ample knowledge about the mechanisms that control the transition from the vegetative to reproductive stages, the continuous identification of new flowering effectors suggests that our knowledge is still fragmented. I aimed in the present study to identify additional uncharacterized genes affecting flowering time and to investigate the effects of additional uncharacterized histone marks on plant genes expression. In combination, this work tries to establish a connection between the different processes involved in the control of gene expression during plant development.

In the first study, I had selected an Arabidopsis mutant with an early flowering phenotype in both long and short day photoperiods. This mutant was further characterized both phenotypically as well as at the genetic and molecular level. Using Illumina deep sequencing, Sanger sequencing and dCAPS markers, the causal mutation was identified and confirmed as a transition from G to A in the BRR2a gene. The brr2a-2 mutation was further confirmed using an allelism test (Chapter 3.1). I found an increased expression of FT and SOC1, moderate decreases in transcripts from MAF genes and significant reduction in FLC transcript levels. These expression changes are consistent with the early flowering phenotype of brr2a-2. The absence of a direct effect on expression of FLC activators or repressors in brr2a-2 suggested an alternative regulatory mechanism. Then, I investigated and confirmed a role of BRR2a in the correct FLC transcript processing. The key splicing control protein BRR2, which is involved in the organization of the spliceosome complex formation in yeast and mammals (Raghunathan and Guthrie, 1998), is evolutionarily conserved also in plants. BRR2a may be required particularly for the splicing of long, highly structured introns, as reflected by the differential decrease in the spliced forms of FLC and other genes’introns.

The identification of nearly 200 different RNA-binding proteins (RBPs) (Lorkovic, 2009) in plants and the potential role of many of them in the FLC transcript processing, [FCA, FY (Simpson et al, 2003); FPA (Schomburg et al, 2001); FLK (Lim et al, 2004); PEPPER (Ripoll et al, 2009); and ELF9 (Song et al, 2009); ABA HYPERSENSITIVE1 (ABH1) (Kuhn et al, 2007); and CstF64, and CstF77 (Liu et al, 2010)], had highlighted the important role of not only transcription but also transcript processing and stability in the control of the floral transition in plants. The identification of BRR2a as a spliceosome component in the control of flowering time widens our understanding of the role of splicing in flowering gene expression regulation. Furthermore, the loading of essential factors for pre-mRNA splicing on nascent transcripts requires phosphorylation of the CTD of Pol II that is established during the transcription elongation process (Kornblihtt et al, 2004; Gornemann et al, 2005; McCracken et al, 1997), suggesting the co-occurrence and cross-communication between splicing and transcription. Defects in spliceosomal components may lead to severe or mild phenotypes depending on the set of affected genes. In human, defects in splicing component may lead to several diseases (Cooper et al, 2009). In Arabidopsis, the complete loss of BRR2a in brr2a-1 insertion mutants leads to seed abortion. In contrast, the BRR2a-T890I protein produced in the brr2a-2 mutant identified here shows only a mild effect on the vegetative and reproductive organ development and affected mainly flowering time.
Taken together, the functional coupling between transcription and pre-mRNA processing to control gene expression is well established, and the identification of a splicing factor in the regulation of flowering time control opens new research areas about splicing in plants. This connects tightly gene expression and splicing control to the floral transition.

The regulation of flowering time gene expression strongly depends on correct chromatin composition and dynamics. Several dozens of distinct histone PTMs were previously identified, and for several of them the roles in establishing active or repressive chromatin states are well defined (Tan et al, 2011, Nakanishi et al, 2009, Liu et al, 2009, Kobza et al, 2005, Fujiki et al, 2011). The complexity of PTMs is not only due to the diversity of these chemical modifications but also to the complex interactions among these marks that we only begin to fully appreciate.

In plants, a regulatory role of many of these PTMs in the control of flowering time as well as expression of FLC and FT is well established (He, 2012). However, little is known compared to the knowledge about histone PTMs in animals. In a second part of this study, I searched for additional uncharacterized histone H3 PTMs in Arabidopsis. Two out of many histone marks not previously identified in plants, were selected for further studies.

H3K23me1 (Chapter 3.3) is a new histone mark associated with repressed genes. H3K23me1 requires KYP, but future work will need to establish whether KYP is directly or indirectly responsible for this modification. H3K23me1 localizes not only on constitutively silent genes, but also on at least some silent Polycomb target genes. This observation suggests a potential colocalization of this mark with H3K27me3 on the same histone tail and a special function in repression of transcription. It is also possible that H3K23me1 recruits LHP1, similar to the recruitment of HP1 by H3K23me1 in animals (Liu et al, 2010d).

I had also identified an interconnection between DNA methylation and H3K23me1, similar to what was observed for H3K9me2 (Du et al, 2012). Such communication between epigenetic marks is the key for proper gene expression control, and future investigations of H3K23me1 function will enlarge our understanding about Polycomb target gene regulation and heterochromatic gene silencing in plants. Similarly, H3K36ac, (Chapter 3.2) an active mark described previously in yeast, ciliates and mammals (Zhang et al, 2012, Morris et al, 2007), was found to be conserved in flowering plants and to be associated with transcribed genes. Unlike in yeast, where H3K36ac localizes at the promoter regions of active genes, in plants this mark is enriched in the transcribed region downstream of the TSS. This distinct localization on gene bodies suggests a different function. This feature is not unique among histone modifications whose effect is often species-dependent. To better understand the role of a given histone modification, it is necessary to study the distribution and functional significance of potentially opposing marks on the same residue, as it was previously established for H3K9ac/me (Shankaranarayana et al, 2003) and for H3K27 ac/me (Wang et al, 2008, Tie et al, 2009). In plants, it has been reported using ChIP-chip that H3K36me3 is highly enriched along the gene body and decreases towards the 3’ end whereas H3K36me2 reaches its peak at the 3’ end. H3K36me2 and H3K36me3 were proposed to repress cryptic promoters in the gene body preventing the expression of non-coding and antisense transcripts. Taken together, it is tempting to speculate that H3K36ac may prevent the expansion of H3K36me2 into promoters and TSS and thus allow efficient transcription. This hypothesis needs to be experimentally tested.

It is important to note that during the transcription process, phosphorylated serine residues on the CTD of Pol II recruit not only pre-mRNA processing proteins (McCracken et al, 1997) but also chromatin-modifying complexes that establish histone marks associated with active genes on the downstream transcribed region. The deposition of H3K36me3 together with H3K4me3 results in the recruitment of histone deacetylases that remove acetylation from coding regions. In addition, chromatin remodelers are
recruited, which favor chromatin opening and sliding of Pol II along the DNA to complete transcription. Taken together, these marks are important in the control of transcription initiation, elongation and termination (Joshi and Struhl, 2005, Wagner and Carpenter, 2012, Li et al, 2007). Furthermore, these processes have been documented well when investigating the regulation of gene expression during plant development. In particular at the flowering transition, trimethylation at H3K4 and H3K36 favors the expression of *FT* and *SOC1* while the enrichment of H3K27me3 on *FLC* establishes a silenced state on this locus and makes it refractory to transcription (Fig. 4.1).

In animals, exons carry often more nucleosomes and more histone PTMs than introns (Kolasinska-Zwierz et al, 2009, Andersson et al, 2009). Some chromatin modifiers and remodelers interact physically with components of the spliceosome machinery as observed in yeast and humans, where the histone acetyltransferase GCN5 interacts with the U2 snRNPs (Gunderson and Johnson, 2009) and where the arginine methyltransferase CARM1 interacts with splicing factors (Cheng et al, 2007). In addition, the CHD1 chromatin remodeler binds to both H3K4me3 and the U2 snRNP (Sims et al, 2007), and the Brahma chromatin remodeler recruits snRNPs U1 and U5 to the pre-transcripts to be spliced (Tyagi et al, 2009). Altogether, this suggests important roles of histone marks not only in the control of transcription but also in the recruitment and correct assembly of pre-spliceosome components on pre-mRNA as well as in the regulation of alternative splicing patterns during physiological processes such as development. Because splicing factors must be loaded onto transcripts early during transcription, it is tempting to speculate that not only H3K4me3 but also H3K36ac, which is highest at the first nucleosome after the TSS, can serve to recruit splicing factors. It will be important to test this possibility in the context of the plant-specific localization of H3K36ac downstream of the TSS.

Action of transcription factors, chromatin modifications and pre-mRNA processing, together form a well-established mechanism that allows precise control of different physiological processes during development. Eventually, deeper understanding of these processes may contribute to the design of targeted drugs against human diseases or to genetically engineered crop plants with optimized time to flower and increased yield.
Figure 4.1. Transcription, splicing and chromatin modifications are connected processes.

Schematic model about the interplay between transcription, splicing and histone modifications in the control of flowering time. In the vegetative stage, FT and FLC are kept silent and actively transcribed, respectively. However, during the transition to flowering, FLC gets silenced while the silencing of FT gets released. The deposition of active histone marks such as H3K36ac and HK4me3 on the amino terminal tail of histone H3 on the nucleosomes located close to the TSS may contribute to the opening of the chromatin structure to allow RNA Pol II (P) initiating transcription. During transcription elongation, additional histone marks are deposited on the nucleosomes along the gene body to allow Pol II sliding and transcription to occur. Simultaneously, the spliceosome protein BRR2a together with the U4 and U6 snRNPs will be loaded onto the pre-mRNA to initiate intron processing. Eventually, the mature transcripts will be released. If the gene is repressed, a compact chromatin structure involving repressive histone marks such as H3K27me3 and possibly H3K23me1 is present.

REFERENCES


130


ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to Prof. Wilhelm Gruissem, the Head of the Plant Biotechnology Department at the ETH Zurich, for excellent conditions and working environment in the laboratory in Zurich, as well as for being a member in my doctoral thesis committee.

I would like to thank Prof. Lars Hennig who granted me the opportunity to work in his group and who helped, and guided me through the doctoral project with his valuable advices.

I would like also to thank Prof. Ernst Hafen, from ETHZ, for being the member of my doctoral thesis committee.

My warmest thanks, I address to all my colleagues in the Prof. Köhler’s and Prof. Hennig’s groups at (ETH, Zurich and SLU, Uppsala) for their openness and support in the practical aspects of the work.

My special thanks go to the previous members of the Plant Biotechnology at ETHZ with whom I had opportunity to interact. I thank all of you for stimulating scientific discussions, technical assistance, and friendly atmosphere during my stay in Zurich.

I should not forget to thank people from the Proteomics Unit at the Functional Genomics Center in Zurich for their valuable technical help. Many thanks as well go to the gardeners in Eschikon, Switzerland as well as at SLU, Sweden, for their careful care of my experiments.

Special thanks go to the members of the Plant Biology and Forest Genetics Department at SLU where I had spent half of my doctoral studies and where I really enjoyed the friendly and scientific atmosphere.

I am very grateful and thankful to my parents and to my wife for their continuous and enormous support.