

Palacký University Olomouc

Diploma thesis

Olomouc 2022

Bc. Jana Vyroubalová

Palacký University Olomouc
Faculty of Science
Department of Cell Biology and Genetics



**Mapping of protein non-coding functional
regions of the barley genome using
epigenetic feature profiling**

Diploma thesis

Bc. Jana Vyroubalová

Field of education: Biologie

Study programme: Molecular and cell biology

Form of study: full-time

Olomouc 2020

Supervisor: Mgr. Pavla Navrátilová, Ph.D.

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2020/2021

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Bc. Jana VYROUBALOVÁ
Osobní číslo: R20917
Studijní program: N1501 Biologie
Studijní obor: Molekulární a buněčná biologie
Téma práce: Mapování protein nekódujících funkčních oblastí genomu ječmene využitím profilování epigenetických znaků
Zadávající katedra: Katedra buněčné biologie a genetiky

Zásady pro vypracování

1. Osvojení si práce s rostlinným materiálem, preparace vybraných tkání ječmene, a příprava chromatinu a lyzátu pro izolaci jader průtokovou cytometrií.
2. Chromatinová imunoprecipitace a ATAC (Assay for Transposase-Accessible Chromatin) s následnou konstrukcí sekvenačních knihoven.
3. Analýza sekvenačních dat s cílem charakterizovat nekódující funkční oblasti a oblasti tkáňově specifické regulace transkripce.

Rozsah pracovní zprávy:

Rozsah grafických prací:

Forma zpracování diplomové práce: tištěná

Jazyk zpracování: Angličtina

Seznam doporučené literatury:


1. Ricci, W. A. et al. Widespread long-range cis-regulatory elements in the maize genome. *Nat Plants* 5, 1237-1249 (2019).
2. Baker, K. et al. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J.* 84, 111-124 (2015).
3. Mascher M. et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544: 427-433 (2017)
4. Lu, Z. et al. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants* 5, 1250-1259 (2019)
5. Yan, W. et al. Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat. Commun.* 10, 1705 (2019)

Vedoucí diplomové práce:

Mgr. Pavla Navrátilová, PhD.
Katedra buněčné biologie a genetiky

Datum zadání diplomové práce: 3. listopadu 2020
Termín odevzdání diplomové práce: 31. července 2021

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA BUNĚČNÉ BIOLOGIE A GENETIKY
L.S. Šlechtitelů 27, 783 71 Olomouc – Holiče
tel.: +420 585 634 901



doc. RNDr. Martin Kubala, Ph.D.
děkan

prof. RNDr. Zdeněk Dvořák, DrSc.
vedoucí katedry

BIBLIOGRAPHICAL IDENTIFICATION

Author's first name and surname	Bc. Jana Vyroubalová
Title	Mapping of protein non-coding functional regions of the barley genome using epigenetic feature profiling
Type of thesis	diploma
Department	Department of Cell Biology and Genetics, Faculty of Science, Palacký University, Olomouc
Supervisor	Mgr. Pavla Navrátilová, Ph.D.
Year of presentation	2022
Keywords	barley, epigenetics, histones, post-translational histone modifications, ChIP-seq, ATAC-seq, enhancers, promoters
Number of pages	70
Number of appendices	0
Language	English

SUMMARY

The theoretical part of this thesis contains a comprehensive literature review incorporating key information on barley plants and epigenetics, with emphasis on enhancers and histone modifications. It also describes how to study epigenetic landscapes, primarily the ATAC-seq and the histone marker ChIP-seq methods. In the experimental part of this work, barley embryos 8 days after pollination (DAP), 24 DAP and 4 days after germination (DAG) seedlings were collected and subjected to ChIP-seq and ATAC-seq analysis. Sequencing data processing and analysis followed using bioinformatical tools.

BIBLIOGRAFICKÉ ÚDAJE

Jméno a příjmení autora	Bc. Jana Vyroubalová
Název práce	Mapování protein nekódujících funkčních oblastí genomu ječmene využitím profilování epigenetických znaků
Typ práce	diplomová
Pracoviště	Katedra buněčné biologie a genetiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci
Vedoucí práce	Mgr. Pavla Navrátilová, Ph.D.
Rok obhajoby práce	2022
Klíčová slova	ječmen, epigenetika, histony, post-translační histonové modifikace, ChIP-seq, ATAC-seq, enhancery, promotory
Počet stran	70
Počet příloh	0
Jazyk	anglický

SOUHRN

Teoretická část této diplomové práce obsahuje literární přehled zahrnující klíčové informace, které se týkají ječmene a epigenetiky, zejména informace o enhancerech a histonových modifikacích. Metody využívané k epigenetickému profilování, především ATAC-seq a ChIP-seq pro post-translační histonové modifikace, jsou taktéž popsány. Experimentální část je zaměřena na studium třech stádií ječmene – embryí 8 a 24 dní po opylení a 4denních klíčků. Tyto tkáně byly podrobeny ATAC-seq a ChIP-seq analýze. Sekvenační data byla zpracována a analyzována za pomoci bioinformatických nástrojů.

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text.

In Olomouc,

Signature

Throughout the writing of this diploma thesis, I have received a great deal of support and assistance. I would first like to thank my supervisor, Mgr. Pavla Navrátilová, Ph.D., whose expertise, feedback and advice were invaluable. I would also like to acknowledge Mgr. Anna Nowicka, Ph.D. for supplying plant material, and Ing. Hana Šimková, CSc. for revision and insight. Another thank you goes to my family and my partner for their patience, help and support. This thesis was part of a project funded by GAČR (21-18794S). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

CONTENTS

1	INTRODUCTION.....	1
2	AIM OF THESIS.....	2
3	LITERATURE REVIEW	3
3.1	Barley (<i>Hordeum vulgare</i>)	3
3.1.1	Barley production and end use.....	3
3.1.2	Academic importance and research	3
3.1.3	Grain development.....	4
3.1.4	Barley genome	6
3.2	Transcriptional regulatory elements.....	6
3.2.1	Promoters	7
3.2.2	Enhancers and silencers	8
3.2.2.1	Characteristics of enhancers	9
3.3	Histone modifications in plants.....	11
3.3.1	Methylation.....	12
3.3.1.1	COMPASS-like complex	13
3.3.1.2	Polycomb repressive complexes (PRCs).....	14
3.3.2	Acetylation.....	15
3.4	Methods and techniques of studying the epigenome	16
3.4.1	Chromatin Immunoprecipitation sequencing.....	16
3.4.2	Assay of Transposase Accessible Chromatin sequencing	18
4	MATERIALS AND METHODS	20
4.1	Biological material	20
4.2	Reagents and solutions	20
4.2.1	Reagents.....	20
4.2.2	Antibodies	20
4.2.3	Buffers.....	20

4.3	Equipment	21
4.4	Methods.....	21
4.4.1	Native ChIP-seq for 8 DAP embryos.....	21
4.4.1.1	Material collection.....	21
4.4.1.2	Chromatin preparation for native ChIP	23
4.4.1.3	Antibody-bead complex preparation and immunoprecipitation.....	24
4.4.1.4	Washes and elution.....	24
4.4.1.5	DNA extraction and library preparation.....	25
4.4.2	Native ChIP-seq for 24 DAP embryos.....	25
4.4.2.1	Material collection.....	25
4.4.2.2	Chromatin preparation 24 DAP for native ChIP	26
4.4.3	Native ChIP-seq for 4 DAG plants	26
4.4.3.1	Material collection and chromatin preparation.....	26
4.4.4	ATAC-seq.....	26
4.4.4.1	Material collection and cross-linking	26
4.4.4.2	Flow-cytometry and nuclei sorting 8 DAP embryos for ATAC-seq..	27
4.4.4.3	Flow-cytometry and nuclei sorting 24 DAP embryos and 4 DAG seedlings for ATAC-seq	27
4.4.4.4	ATAC-seq.....	27
4.4.5	Sequencing.....	28
4.4.6	Data analysis	28
4.4.6.1	ChIP-seq	29
4.4.6.2	ATAC-seq.....	30
5	RESULTS.....	33
6	DISCUSSION.....	44
7	CONCLUSION	46
8	REFERENCES	47

ABBREVIATIONS

ATAC-seq	Assay for transposase accessible chromatin sequencing
BACs	a bacterial artificial chromosome
ChIP-seq	Chromatin immunoprecipitation sequencing
CLF	CURLY LEAF
CTCF	CCCTC-binding factor
DAG	days after germination
DAP	days after pollination
DNase-seq	DNase I hypersensitive sites sequencing
EMF2	EMBRYONIC FLOWER 2
eRNA	enhancer RNA
ESC	EXTRA SEX COMBS
E(z)	ENHANCER OF ZESTE
FAIRE-seq	Formaldehyde-assisted isolation of regulatory elements sequencing
FIS2	FERTILISATION INDEPENDENT SEED 2
FWA	FLOWERING WAGENINGEN
FT	FLOWERING LOCUS T
H3K4me1	mono-methylation at the 4 th lysine residue of the histone H3 protein
H3K4me3	tri-methylation at the 4 th lysine residue of the histone H3 protein
H3K9ac	acetylation at the 9 th lysine residue of the histone H3 protein
H3K27ac	acetylation at the 27 th lysine residue of the histone H3 protein

H3K27me3	tri-methylation at the 27 th lysine residue of the histone H3 protein
HAT(s)	histone acetyltransferase(s)
HDAC(s)	histone deacetylase(s)
HMG-box	high mobility group box
HY5	ELONGATED HYPOCOTYL 5
HYH	ELONGATED HYPOCOTYL 5-Homolog
IGV	The integrative genomics viewer
INR	initiator element
JHDMs	JmjC domain-containing histone demethylases
JmjC	jumonji C
LSD1	Lysine-Specific Demethylase1
LTR	long terminal repeat
MEA	MEDEA
MNase	micrococcal nuclease enzyme
MSI1	MULTIPLE SUPPRESSOR OF IRA 1
N-ChIP	native chromatin immunoprecipitation
NDR	nucleosome-depleted region
PCA	principal component analysis
PcG	Polycomb group proteins
PRC(s)	Polycomb repressive complex(es)
PSAF	PHOTOSYSTEM I SUBUNIT F
RNAPII	RNA polymerase II
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax

SWN	SWINGER
TAD(s)	topologically associated domains(s)
TBP	TATA binding protein
TE(s)	transposable element(s)
TF(s)	transcription factor(s)
TMM	TOO MANY MOUTHS
TSS	transcription start site
VRN2	VERNALIZATION
X-ChIP	cross-linked chromatin immunoprecipitation

LIST OF FIGURES

Figure 1: Phenotype of developing barley seeds of cv. “Compana”. Grain developmental series from 0 (ovary) to 48 DAP (days after pollination), same in dry seeds. Scale bar = 5 mm. Yellow insets show early-stage embryos with a scale bar = 500 μ m (Nowicka *et al.*, 2020). *Stage I:* From 0 to 4 DAP seeds increase in size and weight due to the growth of seed maternal tissues. *Stage II:* From 6 DAP, endosperm increases in size and seed growth accelerates. From 12 to 28 DAP size increases mainly due to the expansion of central starchy endosperm. At 32 DAP the embryo reaches maximum size. *Stage III:* Subsequently, the seeds start desiccating (dry seeds). (Nowicka *et al.*, 2021)..... 5

Figure 2: (A) Dorsal view of barley grain with hulls (glumellae) and after manual removal of the lemma and palea. (B) On the right, a longitudinal cross-section of a barley grain indicating the main structures (Rodríguez *et al.*, 2015)..... 6

Figure 3: Positions of core promoter, proximal promoter, and distal regulatory elements in respect to one another. Core promoter consists of the TSS (+1) and a set of consensus DNA elements, e. g. TATA box and INR element. Near the core promoter TFs can also bind to the proximal promoter. Distal regulatory elements such as silencers and enhancers (see Chapter 3.2.2) located further from the promoter site activate (+) or repress (–) transcription. 8

Figure 4: (A) Enhancer location with respect to the target gene (promoter). (B) Active enhancers in interaction with promoters of their gene through protein complexes. (C) Inactive enhancers and their association with H3K27me3 and H3K4me1. (D) Active enhancers in association with nucleosome-depleted regions and H3K4me1, H3K9ac and H3K27ac (Weber *et al.*, 2016)..... 9

Figure 5: A model of nucleosome assembly from histone dimers. DNA assembled into a nucleosomal structure is depicted at the bottom (Paro *et al.*, 2021). 11

Figure 6: Mono-, di- and tri- methylation are deposited by dynamic lysine methylation catalysed by HKMTs (histone lysine methyltransferases) and removed by histone demethylases (LSD1, lysine-specific demethylase1; JHDMs, JmjC domain-containing histone demethylases) (Liu *et al.*, 2010)..... 13

Figure 7: Histone acetylation and deacetylation gene expression, catalysed by histone acetyltransferases (HATs) and histone deacetylases (HDACs) respectively. Histone acetylation makes chromatin more accessible (“open”, gene activation) and deacetylation is repressive in nature (chromatin is less accessible, gene repression) (Liu *et al.*, 2016). 16

Figure 8: Comparison of experimental protocols. (a) Chromatin immunoprecipitation followed by sequencing (ChIP-seq) for DNA-binding proteins such as transcription factors (TFs). Sample fragmentation by sonication (in the case of ChIP-exo, exonuclease is used for degradation of unbound DNA for more precise localisation of peaks). (b) ChIP-seq for histone modifications. Micrococcal nuclease (MNase) is used to fragment DNA (Furey, 2012).....	18
Figure 9: ATAC-seq relies on the activity of Tn5 transposase, which has the ability to simultaneously cut DNA and insert sequencing adaptors into regions with open chromatin.	19
Figure 10: Insulin needles (A) and tweezers (B) used for embryo dissection.	22
Figure 11: (A) A mature spike (a) and a single spikelet (b) of barley ready for 8 DAP embryo excision. (B) Dissection of a barley seed. The yellow arrow points at the first incision made by a needle/tweezers. The pericarp is removed from the seed. The embryo is situated near the very end of the seed (nearest to the rachis in the spikelet, yellow circle).	23
Figure 12: (A) A mature spike (a) and a single spikelet (b) of barley ready for 24 DAP embryo excision. (B) 24 DAP barley embryo.	25
Figure 13: 4 DAG seedlings – (A) with protective hulls that are removed for the experiment (B)	26
Figure 14: Data analysis workflow – individual steps taken (e.g. trimming), tools used for data adjustment (e.g. TrimGalore) and data file formats (e.g. .fastq).....	28
Figure 15: Visualisation of Bedtools option ‘intersect’.....	30
Figure 16: ATAC-seq data correlation heatmap, Pearson method. The correlation coefficient for samples describes the similarity between them (1.00 being the same sample, 0.0 having no correlation at all). This correlation most importantly describes the similarities and differences between replicas.	35
Figure 17: Principal component analysis (PCA) of ATAC-seq shows replica correlation with eigenvalues of the top two principal components.....	35
Figure 18: Heatmap of ATAC-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). All TSSes are aligned in the middle of the figure, 3 kb around the TSS are displayed. On top of the heatmap, a mean signal around the TSS is shown. Coverage is usually higher on the left side of the TSSes, which signals a promoter region.	36

Figure 19: Correlation of mapped ChIP-seq reads, Pearson method. For ChIP-seq data – three tissue types (8 DAP, 24 DAP, 4 DAG), two replicas each, three histone modifications (H3K4me3, H3K9ac, H3K27me3). The correlation coefficient for samples describes the similarity between them (1.00 being the same sample, 0.0 having no correlation at all). 37

Figure 20: Heatmap of 8 DAP embryos ChIP-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 1 signifies expressed, active genes with high H3K4me3 and H3K9ac; cluster 2 shows unexpressed genes with high H3K27me3 enrichment. 39

Figure 21: Heatmap of 24 DAP embryos ChIP-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 1 signifies expressed, active genes with high H3K4me3 and H3K9ac; cluster 2 shows unexpressed genes with high H3K27me3 enrichment and H3K4me3 enrichment. 40

Figure 22: Heatmap of 4 DAG ChIP-seq around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 2 shows unexpressed genes with high H3K27me3 enrichment and H3K4me3 enrichment; cluster 1 expressed, active genes with high H3K4me3 and H3K9ac. 41

Figure 23: The IGV browser was used to display the epigenomic profiles across the whole barley genome. This specific region of chromosome 3 (chr3H:606 653 004-606 688 204) shows ChIP-seq coverage for H3K9ac, H3K27me3 and H3K4me3 and ATAC-seq peaks and coverage for three different types of tissues, 8 DAP, 24 DAP and 4 DAG. Along with ChIP-seq and ATAC-seq data sets, RNA-seq data sets were integrated, representing gene expression. HORVU.MOREX.r3.3HG0324810.1 gene is located in a repressed region of the genome – high H3K27me3 and low H3K4me3 with H3K9ac, as well as low to none ATAC-seq peaks. HORVU.MOREX.r3.3HG0324830.1 gene on the other hand seems to be fully expressed as it shows high enrichment for H3K4me3, H3K8ac and ATAC-seq

peaks for active chromatin along with peaks for the expressed region from RNA-seq data sets. Another active chromatin region downstream of gene HORVU.MOREX.r3.3HG0324830.1 could possibly be an enhancer region since it does not show signs of expression. 42

LIST OF TABLES

Table 1: Number of ATAC-seq peaks for three different barley tissue types (8 DAP, 24 DAP, 4 DAG). Results of intersection (or pooling in case of 4 DAG) of the two replicates using Bedtools.....	37
Table 2: Number of ChIP-seq peaks for three histone modifications (H3K9ac, H3K4me3, H3K27me3), three different barley tissue types (8 DAP, 24 DAP, 4 DAG). Two replicas each, the intersection of these replicas (Bedtools intersect).	38
Table 3: Data intersection for H3K4me3 and H3K9ac; ATAC-seq, H3K4me3 and H3K9ac; ATAC-seq with genes +1000bp upstream; ATAC-seq with no overlap with genes and promoters (intergenic).....	42
Table 4: Number of intergenic enhancer candidates for three types of barley tissue studied (8 DAP, 24 DAP, 4 DAG).	43

1 INTRODUCTION

Barley is among wheat, rice and maize one of the most essential crops on our planet. It is mainly used as a food source for animals, although it can be seen in the human diet as well. Another valuable use of barley is in malting, in the beer and whiskey industry. This has led to an increasing knowledge of barley grain structure, development and germination. Scientific interest in barley is further perpetuated by the fact, that it is a diploid relative of hexaploid and tetraploid wheats and thus serves as a cereal crop genetic model.

Genome-wide sequencing methods are currently being used to study the epigenome of various organisms, barley included. Using epigenetic feature profiling techniques to map protein non-coding functional regions is an approach that can be considered to be rather new and constantly in development. Among such functional regions are for example enhancer regions, that can activate or repress chromatin and therefore take part in transcription regulation. Methods that can uncover these regions include Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) or Chromatin Immunoprecipitation sequencing (ChIP-seq) among others.

In the experimental part of this thesis, barley embryos 8 days after pollination (DAP), 24 DAP and 4 days after germination (DAG) seedlings were collected and subjected to ChIP-seq and ATAC-seq analysis. Sequencing data processing and analysis followed using bioinformatical tools.

2 AIM OF THESIS

The aim of the theoretical part of this thesis is a comprehensive literature review incorporating key information regarding barley plants, epigenetics (enhancers and histone modifications), as well as methods that are used to study epigenetic landscapes.

The experimental part of this thesis aims to:

- Contribute to deciphering transcriptional regulation of barley genes from their promoter epigenetic signatures.
- Identify putative long-range transcriptional regulatory sequences.
- Define epigenetic changes in promoters and enhancers between three barley developmental stages – 8 DAP, 24 DAP and 4 DAG.

3 LITERATURE REVIEW

3.1 Barley (*Hordeum vulgare*)

3.1.1 Barley production and end use

Barley (*Hordeum vulgare*) is a species of monocot in the family Poaceae (grass family). It ranks fourth among grain cereals (Poaceae species) – maize (*Zea mays*), wheat (*Triticum aestivum*) and rice (*Oryza sativa*) – in terms of production (Sato, 2020). It is grown on both the southern and the northern hemisphere, mostly in temperate regions. Around 65% of global production is accounted for by Europe and the Russian Federation, although it remains an important crop in North Africa, Asia, and South America as well (Langridge, 2018).

Although barley was originally domesticated as a food source intended for humans and has remained an important part of the diet in many regions, food consumption accounts for only 15% of barley end-use. Since the 1960s, the major use of barley has been animal feed, this area takes up between 61-77% of barley use. Another high-value use of barley is malting, this section accounts for about 9-22%. Using barley for malting purposes has led to barley grain structure, development and germination being extensively studied, making it one of the best-studied grains (Schulte *et al.*, 2009).

Despite barley not being used as a food source as extensively, it is considered one of the healthiest cereals in the human diet. This is due to barley containing high amounts of important nutrients. For instance, compared to wheat, it is quite rich in fibre, calcium, iron and zinc. However, it is important to note, that different accessions vary considerably (Langridge, 2018). Barley also contains rather high amounts of β -glucan, a cell-wall polysaccharide that has been found to have a cholesterol-lowering effect. Another advantage of growing barley is its robustness and local adaptability to land, meaning it can be grown in regions where other crop species would fail (Newton *et al.*, 2011).

3.1.2 Academic importance and research

Until recently the main purpose of barley was to serve as a model for wheat. As resources and technologies advanced, it became possible to study wheat independently. However, as stated above, barley is still a very important crop that demands research, especially concerning grain development and germination (Langridge, 2018).

Since barley is also thought to be the first domesticated crop, the study of developmental biology, as well as physiology and biochemistry has been crucial in our understanding of the evolution of crops, but also of humankind. For example, wild barley (*Hordeum vulgare* ssp. *spontaneum*) and the study of its close relationship to the domesticated counterpart has led to the reanalysis of the concept of single-origin of all the major crops along with their domestication (Pourkheirandish *et al.*, 2015; Langridge, 2018). It is believed barley was not domesticated only once, but at least twice, meaning the development of all crops has been more diverse than was previously thought (Morell *et Clegg*, 2007; Langridge, 2018).

Another example of the broad impact barley research could have, is in the field of disease resistance. In this case, the discovery and use (cloning and mutation) of the *Mlo* gene is thought to be a ‘universal weapon’ to combat powdery mildew disease, especially in wheat. This proved to be quite an important step in fighting this disease since powdery mildew disease is caused by over 650 fungal species and invades up to 10 000 plant species (Kusch *et Panstruga*, 2017; Wang *et al.*, 2014). Attempts at breeding mildew-resistant plants usually resulted in growth defects of modified plants in the past. Recently, a successful mutant of mildew-resistant wheat, *Tamlo-R32* variant, without growth or yield penalties has been bred (Li *et al.*, 2022).

3.1.3 Grain development

An essential process in the life cycle of plants, specifically crops, is grain development. Not only are seeds critical for colonisation of the environment, but as stated, grains provide a food source for animals and humans alike. This process is quite complex, involves a multitude of metabolic regulation pathways and is generally divided into three main stages (Weber *et al.*, 2005; Nowicka *et al.*, 2020). Double fertilisation initiates *Stage I* and is characterised by mitotic cell proliferation and little weight gain. In contrast, *Stage II* involves cells differentiation into main tissue types, a large increase in weight, and rapid growth reflecting the accumulation of storage compounds. *Stage III*, on the other hand, is defined by seed maturation, but also weight reduction (desiccation) and finally dormancy (Nowicka *et al.*, 2020). Figure 1 shows weight gain (*Stage I-II*) and loss (*Stage III*) in the first 48 days after pollination (DAP), as well as desiccation of the freshly formed seed.

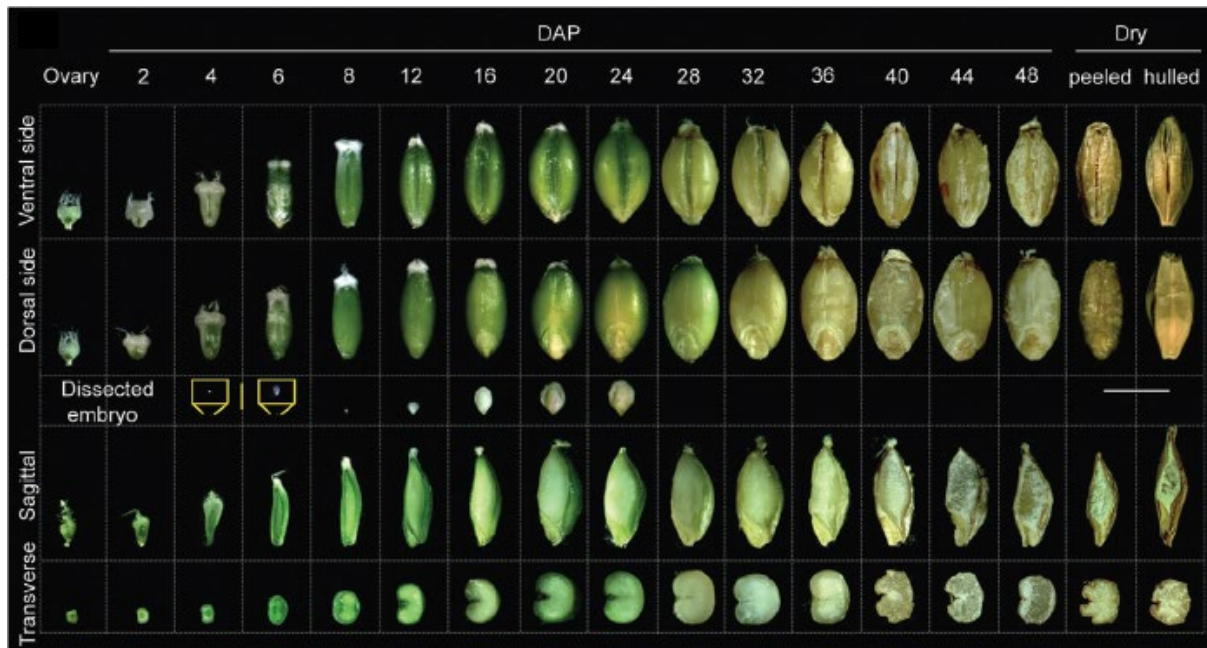


Figure 1: Phenotype of developing barley seeds of cv. “Compana”. Grain developmental series from 0 (ovary) to 48 DAP (days after pollination), same in dry seeds. Scale bar = 5 mm. Yellow insets show early-stage embryos with a scale bar = 500 μm (Nowicka *et al.*, 2020). *Stage I:* From 0 to 4 DAP seeds increase in size and weight due to the growth of seed maternal tissues. *Stage II:* From 6 DAP, endosperm increases in size and seed growth accelerates. From 12 to 28 DAP size increases mainly due to the expansion of central starchy endosperm. At 32 DAP the embryo reaches maximum size. *Stage III:* Subsequently, the seeds start desiccating (dry seeds). (Nowicka *et al.*, 2021).

Diploid embryonic tissues are of maternal and paternal origin. Embryo cell proliferation and differentiation lead to the formation of an embryonic root, shoot apical meristem, cotyledon and plumule (Figure 2). Endosperm (triploid, $3x$) is formed by fertilisation of a diploid central cell and a haploid sperm nucleus (Nowicka *et al.*, 2020). At first a syncytium is formed by the endosperm (the nucleus is pushed to the cell periphery by a central vacuole). Following this process is the formation of a radial network around the nuclei (formed by microtubules), the anticlinal cell wall formation then marks the onset of cell differentiation into endosperm (starch) and the aleurone layer (Olsen, 2001; Nowicka *et al.*, 2020). The mature barley endosperm comprises the central starchy endosperm, aleurone layer, the subaleurone layer, the basal endosperm transfer layer, and the embryo-surrounding region (Olsen, 2001; Sabelli *et al.*, 2009; Nowicka *et al.*, 2020).

Cover of the grain provide seed coats of maternal origin and pericarp. These tissues contain a high amount of starch due to them acting as sustenance, but they also serve a protective function and partake in photosynthesis (Sreenivasulu *et al.*, 2010; Nowicka *et al.*, 2020).

The barley seed itself is protected by diploid hulls (Figure 2), also of maternal origin, which, even after ripening, remain tightly attached to the grain (Rodríguez *et al.*, 2015; Nowicka *et al.*, 2020).

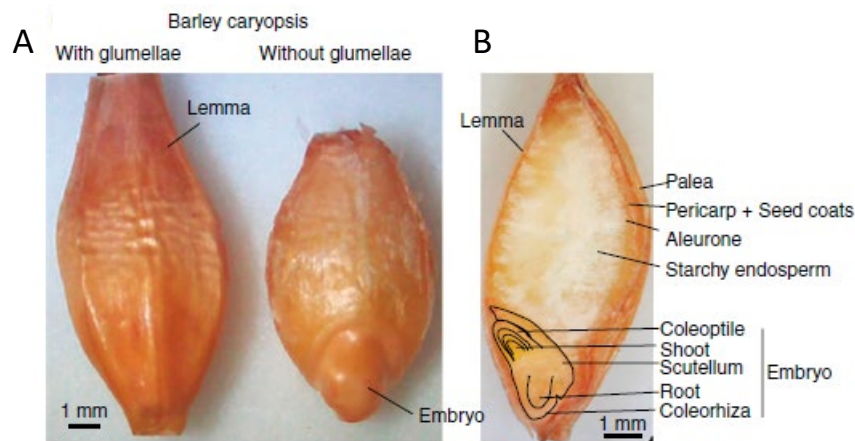


Figure 2: (A) Dorsal view of barley grain with hulls (glumellae) and after manual removal of the lemma and palea. (B) On the right, a longitudinal cross-section of a barley grain indicating the main structures (Rodríguez *et al.*, 2015).

3.1.4 Barley genome

Barley is self-pollinating with a diploid genome consisting of seven chromosomes ($2n = 2x = 14$). It is one of the largest diploid genomes sequenced to date. The latest assembled barley reference sequence is of the cultivar ‘Morex’, a six-row malting variety – MorexV3. The estimation of barley genome size varies greatly among publications, depending on the method, ranging between 3,5 and 5,3 gigabases (Gb). Recently the estimation is thought to be 4.6 Gb for the barley haploid genome with a certain degree of uncertainty attributed to the ribosomal DNA, centromeric and telomeric repeats (Mascher *et al.*, 2021; Sato, 2020; Navrátilová *et al.*, 2022). Approximately 84% of the genome consists of repetitive elements. The majority (76% in random BACs) of these were identified as retrotransposons, most of which (99.6%) are long terminal repeat (LTR) retrotransposons (The International Barley Genome Sequencing Consortium, 2012).

3.2 Transcriptional regulatory elements

Higher eukaryotes consist of many different cell types, with various phenotypes and cellular functions. These cells possess the same DNA, yet there is an astounding diversity among them,

caused by differential regulation of gene expression, spatial and temporal, in response to developmental cues and external biotic and abiotic signals (light, temperature, nutrients, pathogens). Activation and repression of *cis*-regulatory elements at a key moment in time and space is a pathway that gives plants and animals the ability to regulate their gene expression (Kolovos *et al.*, 2012).

The usual definitions of activating regulatory elements focus on two distinct classes – promoters and enhancers. The former defines a group of elements where transcription is initiated and the latter elements that amplify such transcription initiation. These definitions are slowly becoming obsolete since these elements have very similar properties and functions making the distinction between the two disputable (Andersson *et Sandelin*, 2020).

3.2.1 Promoters

For transcription to be initiated, RNA polymerase II (RNAPII) binds to the transcription start site (TSS) defined as the first transcribed genomic nucleotide of a transcript. This binding process is enabled by essential basal transcription factors (TFs) which along with RNAPII form the pre-initiation complex. These TFs usually bind to a specific DNA sequence, typically ± 50 bp around TSS – the ‘core promoter’ (Figure 3). The core promoter determines the precise location of TSS and the direction in which transcription will proceed (Andersson *et Sandelin*, 2020).

The best-known eukaryotic core promoter elements are the TATA-box and the initiator (INR). In the past it was believed that the TATA-box was a universal element, 24-30 bp upstream of TSS and recognised by the TATA-binding protein (TBP), a part of the pre-initiation complex. However, only a small fraction of core promoters found in mammals have a clearly defined TATA-box. Although core promoters are important for the recognition of TSS and the start of transcription, they do not control the temporal and spatial specificity, rate of transcription and elongation. This modulation is caused by TF binding either proximally – to a so-called ‘proximal promoter’, or distally (Figure 3) and aided by recruited co-activators (Andersson *et Sandelin*, 2020).

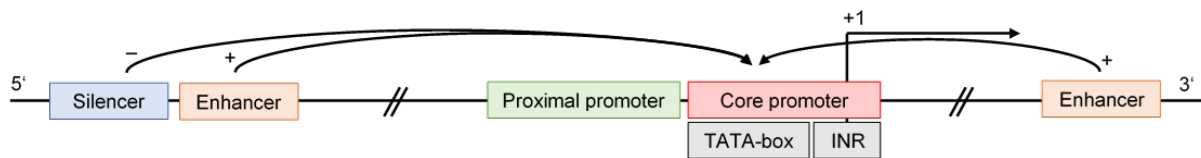


Figure 3: Positions of core promoter, proximal promoter, and distal regulatory elements in respect to one another. Core promoter consists of the TSS (+1) and a set of consensus DNA elements, e. g. TATA box and INR element. Near the core promoter TFs can also bind to the proximal promoter. Distal regulatory elements such as silencers and enhancers (see Chapter 3.2.2) located further from the promoter site activate (+) or repress (-) transcription.

3.2.2 Enhancers and silencers

As mentioned above, transcriptional regulation may also be achieved by a more distal binding of TFs to non-coding DNA sequences relative to promoter regions. Two types of elements are recognized, transcriptional enhancers and silencers, depending on whether they activate or repress target gene expression (Figure 3). These elements can be located up to several Mb away (Weber *et al.*, 2016), up- or downstream of their target genes (i.e., in an orientation-independent manner) or inside of their own or other gene's introns (Ogbourne *et al.*, 1998; Borsari *et al.*, 2021). Important to note is that silencers and enhancers can also be combined into one DNA element (Weber *et al.*, 2016).

Generally, enhancers are activated by binding TFs. Recruitment of co-activators such as histone acetyltransferase and chromatin remodellers follow this process. Together they increase the accessibility of chromatin in a given area (Iwafuchi-Doi *et al.*, 2014). Once the accessibility increases, other TFs can bind which leads to the RNAPII release at the target genes (Shlyueva *et al.*, 2014). Enhancers therefore physically interact with the promoters of their target genes (Figure 4).

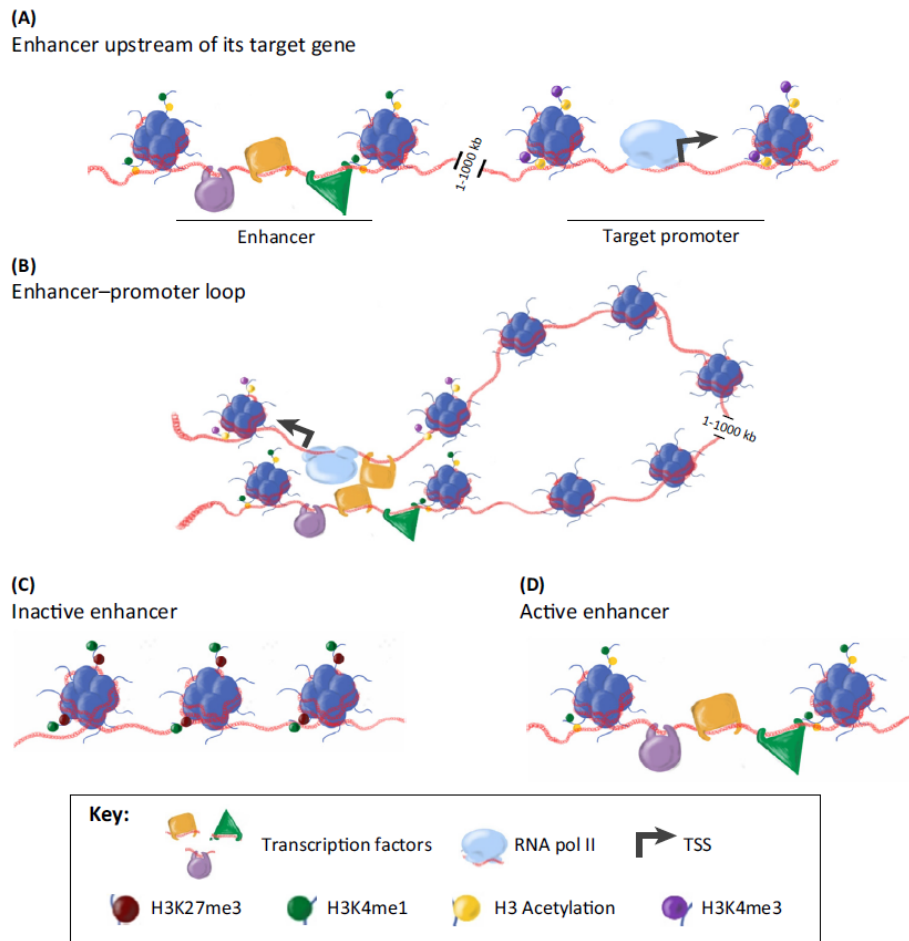


Figure 4: (A) Enhancer location with respect to the target gene (promoter). (B) Active enhancers in interaction with promoters of their gene through protein complexes. (C) Inactive enhancers and their association with H3K27me3 and H3K4me1. (D) Active enhancers in association with nucleosome-depleted regions and H3K4me1, H3K9ac and H3K27ac (Weber *et al.*, 2016).

3.2.2.1 Characteristics of enhancers

In general, several specific characteristics can be identified that enhancer regions display, which help to determine their position in a genome. These include the presence of TF binding motifs, chromatin accessibility, particular histone modifications, eRNA expression, low DNA methylation and physical interactions with their target genes (Shlyueva *et al.*, 2014). Since the study of enhancers is such a complex issue, several methods are usually used in parallel (see Chapter 3.4).

Inactive enhancers are generally displayed by low chromatin accessibility and specific histone modifications, such as H3K27me3. Active enhancers are characterised by high chromatin accessibility and histone acetylation, enhancer RNA (eRNA) transcription and low DNA

methylation (Weber *et al.*, 2016).

TF binding motif

Enhancers are activated by the binding of TFs to a specific DNA sequence bound by the protein's DNA-binding domain (e.g., Homeo-, Zinc-finger or HMG-box). Enhancers can be enriched by several TF binding motifs at once indicating simultaneous co-binding of several TFs. For example, in *Arabidopsis*, more than 530 of these motifs were experimentally identified (Weber *et al.*, 2016).

Chromatin accessibility

Chromatin accessibility depends on the local nucleosome occupancy and binding of chromatin-associated proteins. Accessible genomic regions, also known as nucleosome-depleted regions (NDRs), comprise cell-specific cis-regulatory elements, such as promoters and enhancers. NDRs have been identified and mapped genome-wide in *Arabidopsis*, maize, rice and many others (Kumar *et al.*, 2016; Weber *et al.*, 2016).

Histone modifications

The role of histone modification is a substantial one, a whole chapter of this thesis is dedicated solely to them (see Chapter 3.3).

Enhancer transcripts

Enhancer transcripts (eRNAs) have also proven indicative of active enhancers. eRNAs are short (<2 kb), non-coding, capped, but mostly non-polyadenylated and unspliced RNAs, which are degraded by exosomes quite fast. Some studies suggest that eRNAs are but a by-product of transcription without functional significance, others suggest a greater role of eRNAs and consider them an active player in the recruitment of TFs or mediating enhancer-promoter interactions (Weber *et al.*, 2016; Tan *et al.*, 2022).

DNA methylation

DNA methylation is associated with transcriptional silencing. If present at enhancer sites, it downregulates the expression of target genes. For instance, in *Arabidopsis*, DNA methylation at regulatory sequences was observed in FLOWERING WAGENINGEN (FWA), TOO MANY MOUTHS (TMM) and FLOWERING LOCUS T (FT) (Weber *et al.*, 2016). Therefore, enhancers are usually hypomethylated, since they contribute to transcriptional activation (Liu *et al.*, 2020).

Chromatin interactions

To allow enhancers to activate transcription, enhancers and target genes must physically interact with each other by forming loops. CTCF, Cohesin, one of the key structural protein complexes, together with the Mediator complex, have been shown to mediate also the enhancer-promoter loop interactions (Weber *et al.*, 2016), while producing eRNA. At the same time, inappropriate interactions must be insulated. Insulation of neighbouring regulatory regions forms TADs (topologically associated domains) in mammals and some other eukaryotes. Also, CTCF co-binding with Myc-associated zinc finger protein (MAZ), was found to be on occasion significant in mediating such interactions and binding to cohesin (Xiao *et al.*, 2021). However, plants lack genes encoding CTCF transcription factors posing a challenge to this model.

Growing evidence suggests that H3K27me3 plays a vital role in the spatial organisation of chromatin in eukaryotes (in association with its role as a transcriptional repressor) under the control of Polycomb repressive complexes (PRCs; more in Chapter 3.3.1.1) (Huang *et al.*, 2021).

3.3 Histone modifications in plants

In eukaryotes, the template of transcription is chromatin, with nucleosomes representing the main scaffold as well as the platform for receiving signals to the DNA. A nucleosome consists of an octamer of four histone proteins (H3, H4, H2A, H2B), assembled from two of each H3/H4 and H2A/H2B dimers (Figure 5) (Luger *et al.*, 1997).

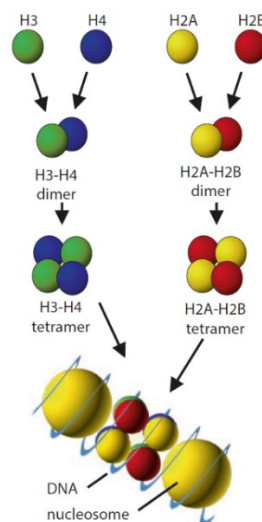


Figure 5: A model of nucleosome assembly from histone dimers. DNA assembled into a nucleosomal structure is depicted at the bottom (Paro *et al.*, 2021).

The state of chromatin (repression or activation) influences accessibility to the transcriptional machinery and is regulated by mechanisms such as DNA methylation, post-translational histone modifications, exchange of histone variants, chromatin remodelling or incorporation of non-coding RNAs. The established patterns can be heritable to successive generations, therefore they are considered epigenetic (Zhao *et al.*, 2019; Varshney *et al.*, 2018).

Histone modifications refer to the post-translational covalent modifications on the N-terminal tails of histones (Duan *et al.* 2018; Zhao *et al.*, 2019). These modifications are a part of ‘histone code’, which specifies the type of modification, its function and transcriptional activity (Jenuwein *et al.*, 2001; Duan *et al.*, 2018). There have been reported over 100 different amino acid residues on histones, the most frequently studied include methylation, acetylation ubiquitination, sumoylation and phosphorylation. Among others are for example propionylation, butyrylation, formylation, citrullination, proline isomerization, and ADP ribosylation. (Tan *et al.*, 2011; Duan *et al.* 2018; Zhao *et al.*, 2019). Epigenetic modifications are species-, tissue-, organelle-, and age-specific, they are involved in various processes such as transposon repression, genomic imprinting and stress-associated defence responses (Varshney *et al.*, 2018; Gehring, 2013).

3.3.1 Methylation

Methylation of histones not only occurs at different residues – lysine (Lys, K) or arginine (Arg, R); it also occurs in different sites (4, 9, 27 etc.) and the number of added methyl groups differs as well (mono-, di- or trimethylation). This variability plays an essential role in multiple biological processes such as transcriptional regulation of transposable elements (TEs) and protein-coding genes during plant development and stress response (Liu *et al.*, 2010; Duan *et al.*, 2018).

Lysine methylations are among the most investigated due to their roles as both transcriptional activators and repressors. In plants, lysine methylation occurs on histone H3 at four sites 4, 9, 27 and 36. For instance, H3K9 and H3K27 methylations are generally considered repressive, mostly associated with silenced regions. In contrast, H3K4 and H3K36 methylations appear in actively expressed genes (Duan *et al.*, 2018).

Lysine methylation is catalysed by a group of SET-domain-containing histone methyltransferases. Removal of this type of methylation by two types of histone demethylases

–jumonji C (JmjC) domain-containing proteins and Lysine-Specific Demethylase1 (LSD1) like proteins (Figure 6) (Liu *et al.*, 2010; Zhao *et al.*, 2019).

As an example, H3K4me3 and histone acetylation are responsible for the active expression of FLOWERING LOCUS (FLC) which results in late flowering. On the other hand, H3K9me2, H3K27me2 and histone deacetylation reverse this effect by repression of FLC in *A. thaliana* (He, 2009).

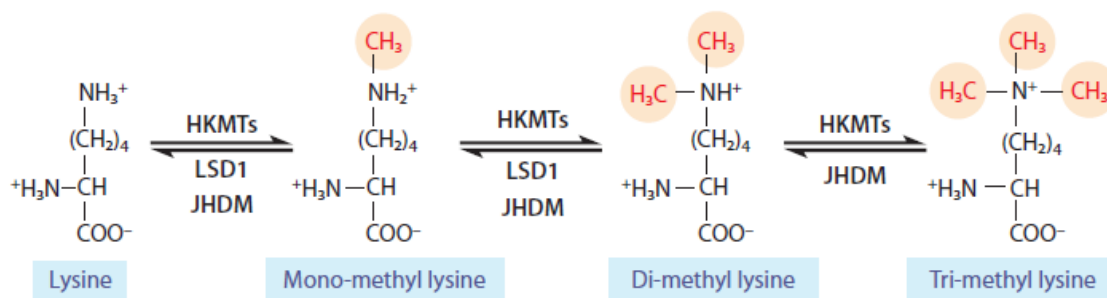


Figure 6: Mono-, di- and tri- methylation are deposited by dynamic lysine methylation catalysed by HKMTs (histone lysine methyltransferases) and removed by histone demethylases (LSD1, lysine-specific demethylase1; JHDMs, JmjC domain-containing histone demethylases) (Liu *et al.*, 2010).

3.3.1.1 COMPASS-like complex

Methylation of lysine 4 of histone H3 (H3K4), especially H3K4me3 is associated with active gene expression. In animals (*Drosophila* and mammals), both di- and trimethylation of H3K4 are related to transcriptional activation. However, in plants (*Arabidopsis*) only H3K4me3 has been implied to be activating. H3K4 methylation is catalysed by the H3K4 methyltransferase complexes termed COMPASS-like complexes. The COMPASS complex was first identified in budding yeasts (*Saccharomyces cerevisiae*) and consists of seven polypeptides (Set1, Cps60, Cps50, Cps40, Cps35, Cps30, and Cps25). In yeasts, histone methyltransferase Set1 carries out H3K4me1/2/3 and requires interaction with other COMPASS subunits to fully function (Jiang *et al.*, 2011; Zhao *et al.* 2022).

Most COMPASS subunits are fully conserved in animals and plants. For instance, in humans, six types of these complexes have been identified. They all contain WD repeat domain 5 (WDR5; homologous to Cps30), retinoblastoma binding protein 5 (RbBP5; homologous to Cps50), and absent, small or homeotic 2-like (ASH2L; homologous to Cps60) proteins, along

with one of several specific methyltransferases (Set1a, Set1b, MLL1, MLL2, MLL3 or MLL4). In *Arabidopsis*, homologs of WDR5, RbBP5 and ASH2L form a subcomplex that interacts with H3K4 methyltransferases to form a COMPASS-like complex. For example, floral transition and plant development are regulated by these complexes (Jiang *et al.*, 2011; Zhao *et al.* 2022)

3.3.1.2 Polycomb repressive complexes (PRCs)

The trimethylation of lysine 27 of histone H3 (H3K27me3) is a histone modification that is generally considered repressive. This is due to the activity of Polycomb repressive complexes (PRCs) that are formed by a combination of multiple Polycomb group (PcG) proteins (Grossniklaus *et al.*, 2014). In both animals and plants, biochemical analyses have shown that PcG proteins normally assemble into two large multiprotein groups with different histone-modifying activities – Polycomb repressive complex 1 (PRC1) and 2 (PRC2). PRC2 deposits H3K27me3 on its targets and PRC1 recognizes this modification and stabilizes PRC2-mediated repression (Huang *et al.*, 2021; Baile *et al.*, 2022). For example, PRC1 has H2A E3 ubiquitin ligase activity on lysine 119, 120 and 121 in *Drosophila*, vertebrates and *Arabidopsis*, respectively (Baile *et al.*, 2022).

The PRC2 complex (also known as EXTRA SEX COMBS-ENHANCER OF ZESTE (ESC-E(Z)) complex) is one of three PcG complexes in *Drosophila*. It has been shown to play a role in epigenetic silencing of target genes involved in cell growth and proliferation and in early development both in *Drosophila* and mammals (Kapalazogou *et al.*, 2010). *Drosophila* PRC2 core is composed of Enhancer of Zeste (E(z); SET-domain containing methyltransferase), Suppressor of Zeste 12 (Su(z)12; scaffold protein), Extra sex combs (Esc; H3K27me3 binding protein) or Esc-like (Esc1) and nucleosome remodelling factor (Nurf55). In plants, different variants of the PRC2 complex have been identified (they differ depending on their developmental stage).

In *Arabidopsis*, homologs of *Drosophila* PRC2 core subunits have been found. CURLY LEAF (CLF), MEDEA (MEA) and SWINGER (SWN) are homologs of E(z). FERTILISATION INDEPENDENT SEED 2 (FIS2), EMBRYONIC FLOWER 2 (EMF2) and VERNALIZATION 2 (VRN2) are homologs of Su(z) 12. FERTILIZATION INDEPENDENT ENDOSPERM (FIE) is a unique Esc homolog and MULTIPLE SUPPRESSOR OF IRA 1 (MSI1) is the homolog of Nurf55. *Arabidopsis* has at least three different PRC2 cores: VRN2-

PRC2 (VRN2, CLF/SWN, FIE, MSI1), SMF2-PRC2 (EMF2, CLF/SWN, FIE, MSI1) and FIS2-PRC2 (FIS2, MEA, FIE, MSI1). These PRC2 cores display distinct but also overlapping functions in the regulation of gene expression in plants (Baile *et al.*, 2022). The best-studied one is the FIE/MEA complex, which regulates the initiation of seed development. It consists of FIE, MEA, FIS2 and MSI1. Moreover, the EMF2-CLF/SWN-FIE-MSI1 complex has been suggested to play a suppressive role in the transition from vegetative development into flowering and flower formation. Another process regulated by VRN2-CLF/SWN-FIE-MSI1 is vernalisation (Kapalazogou *et al.*, 2010).

It is becoming increasingly evident that the spatial organisation of chromatin in eukaryotes is under the influence of H3K27me₃. For instance, in animals, Polycomb targets marked with this histone modification establish interaction between them and thus form a repressive chromatin hub that is fully dependent on PRCs and insulator proteins. In mammalian embryonic stem cells, PRCs have been shown to regulate the maintenance of pluripotency. In *Drosophila*, high levels of H3K27me₃ are associated with genomes subdivision into TADs. In plants, studies on this topic are few, but some of them suggest this specific covalent histone modification could be the key contributor to chromatin topology (Huang *et al.*, 2021).

3.3.2 Acetylation

Acetylation can be found at many lysine residues of histone H2A, H2B, H3 and H4. Modulation of acetylation is procured by histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Figure 7) (Zhao *et al.*, 2019). The presence of histone acetylation leads to the neutralization of charges of lysine residues. Thus, the interaction between histone and DNA are weakened and chromatin becomes more accessible to regulators and allows active transcription and vice versa (Jiang *et al.*, 2020). N-terminal lysine residues of histone H3 can be acetylated at positions 9, 14, 18, 23 or 27; on histone H4 at positions 5, 8, 12, 16 or 20 (Liu *et al.*, 2016).

For example, histone modification H3K9ac contributes to light-induced activation of ELONGATED HYPOCOTYL 5 (HY5) and ELONGATED HYPOCOTYL 5-Homolog (HYH) and their downstream effectors like photosynthesis-related genes such as PHOTOSYSTEM I SUBUNIT F (PSAF) (Charron *et al.*, 2009).

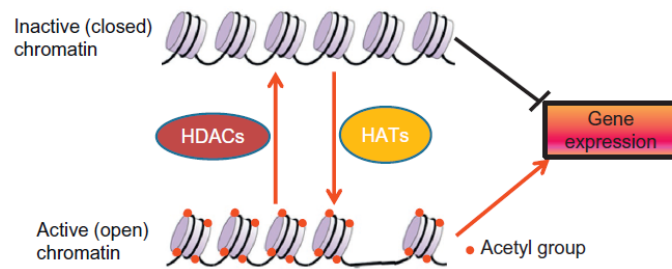


Figure 7: Histone acetylation and deacetylation gene expression, catalysed by histone acetyltransferases (HATs) and histone deacetylases (HDACs) respectively. Histone acetylation makes chromatin more accessible (“open”, gene activation) and deacetylation is repressive in nature (chromatin is less accessible, gene repression) (Liu *et al.*, 2016).

3.4 Methods and techniques of studying the epigenome

Genome-wide sequencing-based methods have been recently developed to address various biological questions that were previously left unanswered due to limited technology. Their application to characterize the epigenetic landscape gave rise to the field of epigenomics (Reske *et al.*, 2020; Yan *et al.*, 2020). Current applications include gene expression, protein-DNA interactions, DNA methylation, histone post-translational modifications, 3D genome organisation, chromatin accessibility and nucleosome occupancy (Reske *et al.*, 2020).

These newly developed techniques are usually applied to multiple cellular states, utilize several methods and are followed by integration and comparative analyses. Methods used to study the epigenome include Assay of Transposase Accessible Chromatin sequencing (ATAC-seq), DNase I hypersensitive sites sequencing (DNase-seq), Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIREseq) – all of these methods investigate chromatin accessibility. Chromatin Immunoprecipitation sequencing (ChIP-seq) on the other hand, is used to identify transcription factor binding sites and histone modifications (Yan *et al.*, 2020). Furthermore, Hi-C is a method developed for the identification of global genome interactions and the depiction of 3D genomic architecture (Kong *et al.*, 2019).

3.4.1 Chromatin Immunoprecipitation sequencing

Chromatin immunoprecipitation (ChIP) is an essential method to isolate DNA fragments bound to specific proteins *in vivo* utilizing antibody-protein interaction. It enables the study of transcription factor binding sites, as well as histone modifications in a genome-wide manner

(Cortijo *et al.*, 2018). In the past, the most common protocol was based on ChIP assay followed by qPCR or hybridisation of isolated DNA fragments to a microarray (also called ChIP-chip). Nowadays, with the rapid development of high-throughput sequencing, the DNA fragments obtained by ChIP are sequenced directly (Chen *et al.*, 2018).

We can differentiate between two types of ChIP-seq – native (N-ChIP) and cross-linked (X-ChIP). In X-ChIP, the plant material is fixed, preferably at the moment of collection, using chemicals such as formaldehyde, which cross-links DNA-bound proteins to the DNA *in vivo*. Lysis of the cells follows, and chromatin is extracted and fragmented. Two main methods are used for fragmentation, sonication and digestion. Sonication results in a fragment of an average size of ~300 bp (ranging between 100 to 1000 bp) (Figure 8a). The average length of DNA fragmented by sonication depends on the duration and strength of sonication. Another method used is digestion with the micrococcal nuclease (MNase) enzyme (Figure 8b). This fragmentation is used mainly on native chromatin and in X-ChIP it can be recommended especially when mono-nucleosome resolution is needed (e.g., to reach high resolution or analysis of nucleosome stability) (Cortijo *et al.*, 2018; Chen *et al.*, 2018; Huang *et al.*, 2020).

In both cases (N- and X-ChIP), fragmented chromatin (protein/DNA complex) is immunoprecipitated by a specific antibody that recognizes the protein of interest, coupled with paramagnetic beads. Washing steps remove proteins or DNA non-specifically bound to the beads before elution of protein/DNA complexes. In the case of X-ChIP, an extra step of reversing the cross-link is needed. By combination of detergent, high temperature and proteinase treatment, the immunoprecipitated DNA is released, followed by purification and further analysis (Cortijo *et al.*, 2018; Chen *et al.*, 2018; Huang *et al.*, 2020). For most non-histone chromosomal proteins, X-ChIP is the only option since nuclease digestion proved to digest also open chromatin regions relatively loosely bound to TFs, which also need to be covalently attached by crosslinking before fragmentation. However, N-ChIP is preferable when profiling histones and histone modifications for better antibody specificity, lower background and mononucleosomal resolution (Huang *et al.*, 2020).

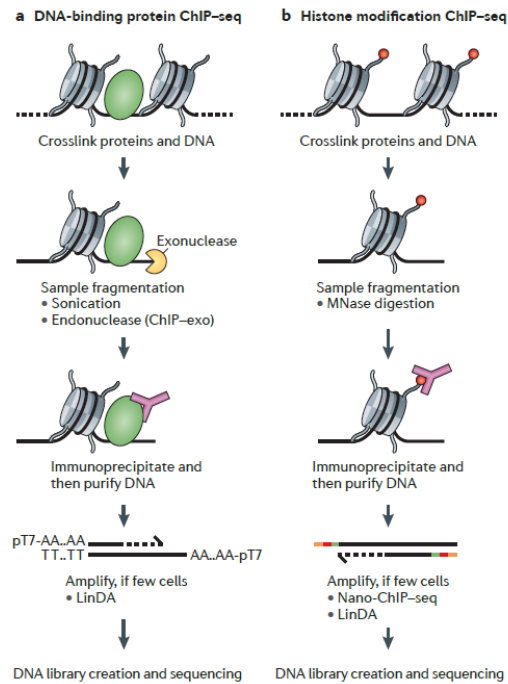


Figure 8: Comparison of experimental protocols. **(a)** Chromatin immunoprecipitation followed by sequencing (ChIP-seq) for DNA-binding proteins such as transcription factors (TFs). Sample fragmentation by sonication (in the case of ChIP-exo, exonuclease is used for degradation of unbound DNA for more precise localisation of peaks). **(b)** ChIP-seq for histone modifications. Micrococcal nuclease (MNase) is used to fragment DNA (Furey, 2012).

3.4.2 Assay of Transposase Accessible Chromatin sequencing

Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) has rapidly become one of the most powerful tools to study chromatin accessibility. ATAC-seq relies on the activity of Tn5 transposase, which simultaneously cuts DNA and inserts sequencing adaptors into regions with open chromatin (Figure 9). Tn5 transposase is a prokaryotic enzyme naturally encoded by the Tn5 transposon, which contains specific 19 bp flanking regions – end sequences (ESs). In its native environment, Tn5 recognises ESs of the transposon and through a cut-and-paste mechanism, it excises the DNA and inserts it into a new position. Science has exploited this ability *in vitro* and utilizes a derivative hyperactive Tn5 transposase bound to a DNA sequence that is to be inserted. More recently, this reaction has been further improved to allow simultaneous production of DNA sequencing libraries, a procedure known as tagmentation, which adds adaptor sequences to DNA fragments. The size distribution of the DNA fragments can be controlled by changing the concentration of transposase complexes relative to the target DNA (Shashikant *et Eppensohn*, 2019).

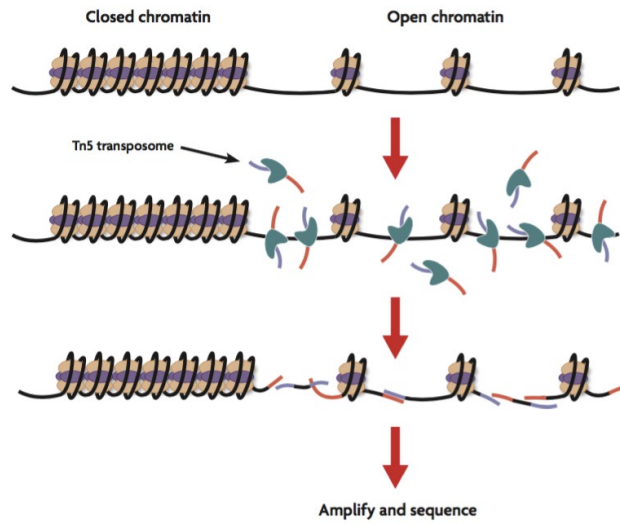


Figure 9: ATAC-seq relies on the activity of Tn5 transposase, which has the ability to simultaneously cut DNA and insert sequencing adaptors into regions with open chromatin.

4 MATERIALS AND METHODS

4.1 Biological material

- Barley cv. “Morex” embryos – 8-10 DAP and 24 DAP
- Barley cv. “Morex” seedlings 4 DAG

4.2 Reagents and solutions

4.2.1 Reagents

- cOmplete™ Protease Inhibitor Cocktail (Roche, #11697498001)
- Micrococcal nuclease (NEB, #M0247S)
- ChIP DNA Clean & Concentrator (Capped Columns) (Zymo Research, #D5205)
- Dynabeads™ Protein G for Immunoprecipitation (Invitrogen, ThermoFisher Scientific, #1003D)
- Phasemaker™ Tubes (Invitrogen, ThermoFisher Scientific, #A33248)
- NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB, #E7645S/L)
- Accel-NGS™ 2S DNA Library Kit for Illumina (Swift Biosciences, #21024)
- ATAC-Seq Kit (Active Motif, #53150)
- NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycles) (Illumina, #20028313)

4.2.2 Antibodies

- Anti-H3K9ac (Millipore, #07-352)
- Anti-H3K4me3 (Millipore, #04-745)
- Anti-H3K27me3 (Diagenode, #C15410195)

4.2.3 Buffers

TC buffer: 50 mM Tris-HCl (pH 7.5), 75 mM NaCl, 6 mM MgCl₂, 0.1 mM CaCl₂, 1x protease inhibitor cocktail (PI)

10x HB buffer (pH 9.4): 40 mM spermidine, 10 mM spermine, 0.1 M Na₂EDTA, 0.1 M Trizma base, 0.8 M KCl.

1x H buffer: 1x HB, 0.5 M sucrose, 0.5% (v/v) Triton X-100, 0.1% (v/v) 2-mercaptoethanol, 1x protease inhibitor cocktail

MNase digestion buffer: 50 mM Tris-HCl (pH 7.5), 0,2% Triton X-100, 10 % sucrose, 4 mM MgCl₂, 1 mM CaCl₂, 1x protease inhibitor cocktail.

ChIP incubation buffer: 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM EDTA, 0.1% SDS (sodium dodecyl sulphate), 0.1% sodium deoxycholate, 1% Triton X-100, 1x protease inhibitor cocktail.

High salt ChIP buffer: 10 mM Tris-HCl (pH 7.5), 300 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 1x protease inhibitor cocktail.

Bead wash buffer: 0.02% Tween-20 in 1x PBS.

First elution buffer: 1% SDS in TE.

Second elution buffer: 500 mM NaCl in TE.

MNase dilution buffer: 5 mM Tris with HCl (pH 7.7), 25 μ M CaCl₂, 50% glycerol.

Lysis buffer LB01: 15mM Tris, 2mM Na₂EDTA, 0.5mM spermine (4HCl), 80mM KCl, 20mM NaCl, 0.1% Triton X-100, pH 9.0; after filtration add 15mM 2-mercaptoethanol.

4.3 Equipment

- Wide Zoom Stereo Microscope SZX16 (Olympus)
- Vacuum system LABOBASE SBC 840 (Labobase)
- Qubit 4 Fluorometer (ThermoFisher)
- Bioanalyzer Instrument 2100 (Agilent)
- 120 IEC-Multi RF Refrigerated Centrifuge (Thermo Scientific)
- Multifuge X1R Centrifuge (Thermo Scientific)
- ThermoMixer C (Eppendorf)
- Dynabeads™ MX Mixer (Invitrogen)
- NovaSeq™ 6000 (Illumina)
- Flow cytometer BD FACSAria™ (SORP)

4.4 Methods

4.4.1 Native ChIP-seq for 8 DAP embryos

4.4.1.1 Material collection

To obtain embryos 8-10 days after pollination, barley plants were examined. Plants about 2 months of age during anthesis have a specific look – spikes having anthers showing but not shedding yet were removed and used to collect caryopses (grains) of about 1 cm in length. The caryopsis was dissected using insulin needles and tweezers under a stereomicroscope to excise

the embryo (Figure 10). First, an incision was made approximately 3 mm from the end, which holds the embryo, and then another was done alongside the caryopsis to the pinnacle (Figure 11). The outer layer (pericarp) was removed using tweezers. By gently pushing from the inside to the pinnacle, the embryo was detached from the rest of the grain. It was then moved into a microtube filled with 1x PBS and put on ice. The embryos in 1x PBS were centrifuged, the supernatant was removed, and the embryos were fixed in 1% formaldehyde in PBS. The sample was put into a vacuum for 8 min. 0.125M glycine was added and the sample was again put into the evaporator for 5 min. The sample was centrifuged and the supernatant replaced with 1x PBS wash buffer repeatedly, but carefully to prevent loss of embryos. The sample was then snap-frozen in liquid nitrogen and kept at -80°C until further use.



Figure 10: Insulin needles (A) and tweezers (B) used for embryo dissection.

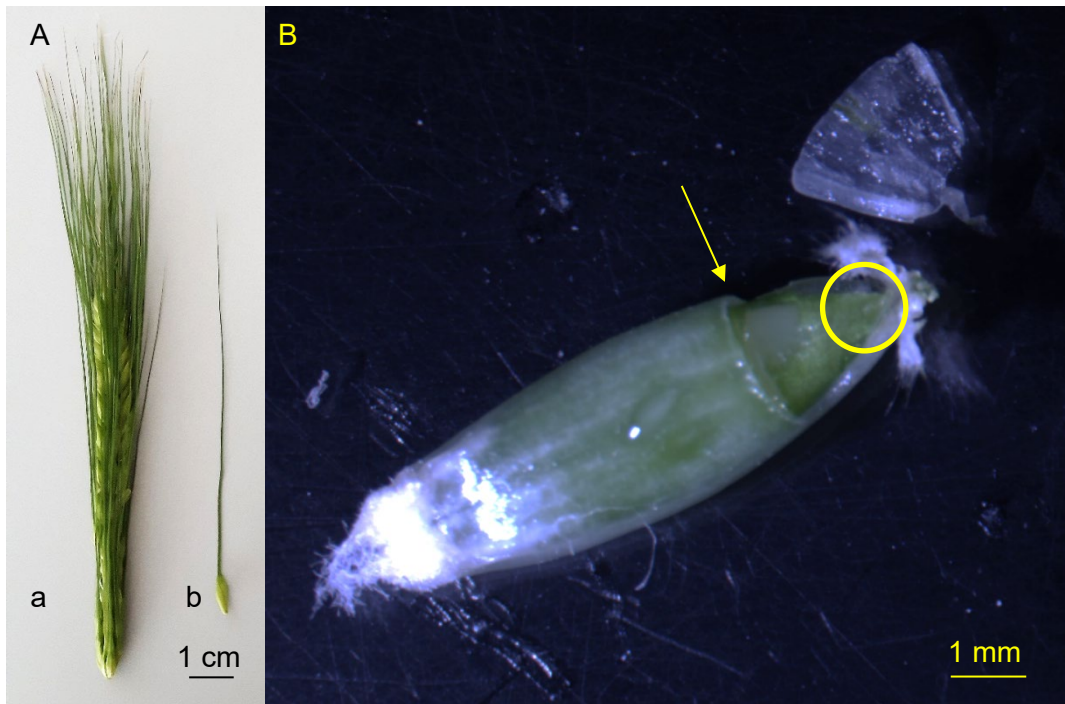


Figure 11: (A) A mature spike (a) and a single spikelet (b) of barley ready for 8 DAP embryo excision. (B) Dissection of a barley seed. The yellow arrow points at the first incision made by a needle/tweezers. The pericarp is removed from the seed. The embryo is situated near the very end of the seed (nearest to the rachis in the spikelet, yellow circle).

4.4.1.2 Chromatin preparation for native ChIP

Approximately 50-60 Barley cv. “Morex” 8-10 DAP embryos per sample were taken and kept on ice. Embryos were resuspended in H-buffer and centrifuged (3000g, 15 min, 4 °C). In order to homogenize the material and release nuclei, Dounce homogenizer or pellet pestle was carefully used. The homogenization was completed by pushing the suspension three times through an insulin needle. The suspension was filtered through 100 µm and 20 µm Cell Strainer and centrifuged (3000g, 15 min, 4 °C). Nuclei were resuspended in fresh H-buffer and samples were centrifuged once more (3000g, 15 min, 4 °C) followed by resuspending in TC-buffer and centrifuged (3000g, 15 min, 4 °C) yet again.

Nuclei were taken up in MNase buffer, approx. 100 µl/sample. Chromatin was digested at 37 °C for 10 min, with varying amounts of MNase to achieve fragments of 1-3 nucleosomes in size.

The volume of MNase added was optimised before first use by chromatin:MNase titration. Chromatin aliquots were digested by 0-2 µl of MNase at 37 °C for 10 min. Aliquots of 20 µl were taken from each digestion, 20 µl of TE with 1% SDS was added with Proteinase K and

proteins were digested for 1 hour at 55 °C. DNA was isolated by AMPure beads and fragment size distribution was analysed. The 2x ChIP incubation buffer was added in equal volume (1:1) to reach immunoprecipitation conditions and stop the MNase reaction. Samples were pooled and centrifuged (13000g, 10 min, 4 °C) to obtain soluble chromatin without particles. The supernatant was taken into a separate microtube and 10% of the volume was set aside to be used as input (1/10th of the volume of one ChIP reaction).

Size distribution of fragments was checked by running a 1.5% agarose gel of isolated DNA or by Bioanalyzer.

4.4.1.3 Antibody-bead complex preparation and immunoprecipitation

For antibody-bead complex preparation, paramagnetic Protein-G Dynabeads with a magnetic stand were used following a manufacturer's manual. Briefly, beads were washed in an excess of Bead Wash buffer, then resuspended in 10x excess of the Bead Wash buffer, antibodies were added (2.5 µg per 10 µl of beads), and bead suspension was incubated (2 h, 4 °C, rotating). Bead-antibody complexes were washed twice with Bead Wash buffer and once with ChIP incubation buffer to remove free antibodies. Chromatin solution was added to the prepared beads and samples were left to rotate for at least 4 hours at 4 °C (usually overnight) to immunoprecipitate. Importantly, protein low-binding tubes were used to prevent unspecific background.

4.4.1.4 Washes and elution

All washes and elution were performed by using a magnetic stand. Beads were washed 3x for 5 min with ChIP incubation buffer, 2x for 5 min with High Salt ChIP buffer and once for 5 min with TE buffer. With the swap to TE buffer, a new low bind microtube was taken.

Elution started with the swap of buffer to 75 µl of First Elution buffer and incubated (850 rpm, 15 min, 55 °C). The supernatant was collected into a new microtube. Second Elution buffer was added to the beads, followed by incubation (850 rpm, 15 min, 55 °C). After the second elution, supernatants from both elution steps were pooled for each sample (antibody type). Proteinase K was added to all samples (3 µl/sample of 20 mg·ml⁻¹) and incubated (55 °C, 1 h). Input samples were prepared similarly (75 µl of First Elution buffer and 75 µl of Second Elution

buffer with Proteinase K).

4.4.1.5 DNA extraction and library preparation

DNA was purified with the ChIP DNA Clean & Concentrator kit into a final amount of 20 μ l. Alternatively, using phenol:chloroform:isoamylalcohol in the Phase-Lock tubes followed by ethanol precipitation is possible. DNA concentration was measured by Qubit.

NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® was used for library preparation. For total DNA amount under 0.5 ng Accel-NGS™ 2S DNA Library Kit for Illumina was used.

4.4.2 Native ChIP-seq for 24 DAP embryos

4.4.2.1 Material collection

The preparation of 24 DAP embryos was similar to 8 DAP. Grains were removed from spikes and embryos were easily removed from them by tweezers (Figure 12). After collecting a desirable number of embryos, the same protocol was followed as with 8 DAP embryos (see Chapter 4.4.1.1).

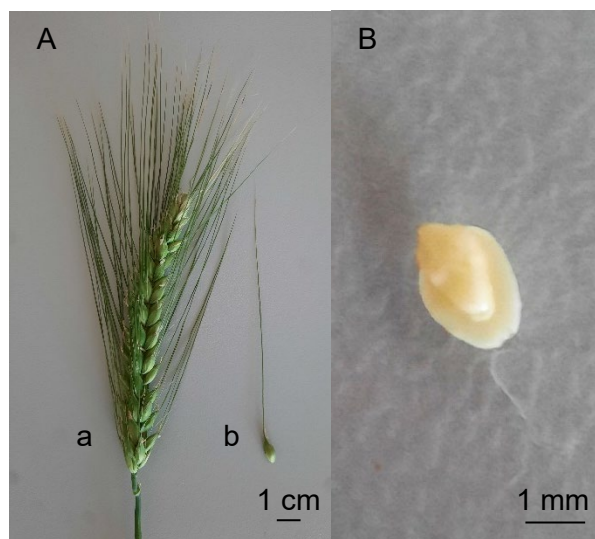


Figure 12: (A) A mature spike (a) and a single spikelet (b) of barley ready for 24 DAP embryo excision. (B) 24 DAP barley embryo.

4.4.2.2 Chromatin preparation 24 DAP for native ChIP

Barley cv. “Morex” 24 DAP embryos were dissected, approximately 10 embryos per sample, and kept on ice. Embryos were ground in a mortar with liquid nitrogen and powder was collected into a tube with H-buffer. The rest proceeded in the same manner as in Chapters 4.4.1.2–4.4.1.5. However, the suspension was filtered through 45 µm Miracloth nylon mesh on a funnel instead of Cell Strainer.

4.4.3 Native ChIP-seq for 4 DAG plants

4.4.3.1 Material collection and chromatin preparation

Approximately 4 g of barley 4 DAG seedlings were taken (Figure 13). Rinsed in water, dried and snap-frozen in liquid nitrogen. Using a pre-cooled mortar and pestle, the tissue was ground. The powder was incubated in 25 ml of H-buffer for 5 minutes on ice and then filtered through 45 µm Miracloth nylon mesh. The protocol proceeded the same as with 8 DAP and 24 DAP embryos, exception being the amount of H-buffer (25 ml) and TC-buffer (10 ml) used for rinsing (see Chapters 4.4.1.1–4.4.1.5).

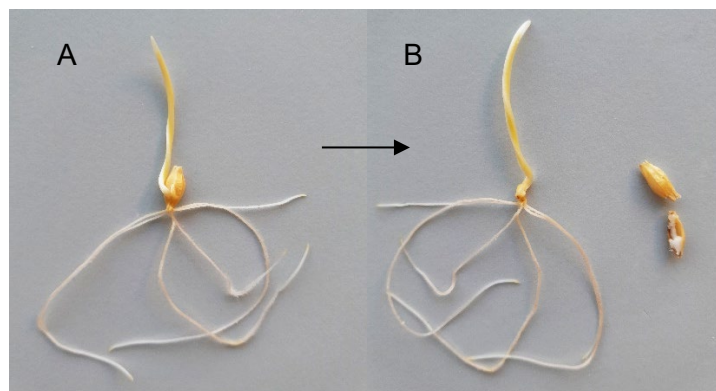


Figure 13: 4 DAG seedlings – (A) with protective hulls that are removed for the experiment (B).

4.4.4 ATAC-seq

4.4.4.1 Material collection and cross-linking

The same procedure was followed as in Chapters 4.4.1.1, 4.4.2.1 and 4.4.3.1 for 8 DAP, 24 DAP and 4 DAG respectively. However, embryos were fixed in 1% methanol-free formaldehyde to cross-link proteins and chromatin and enable flow-sorting.

4.4.4.2 Flow-cytometry and nuclei sorting 8 DAP embryos for ATAC-seq

Approximately 50-60 Barley cv. “Morex” 8 DAP fresh embryos per sample were taken up in LB01 buffer with 1x protease inhibitor and centrifuged (3000g, 15 min, 4 °C). In order to homogenize the material and release nuclei, Dounce homogenizer or pellet pestle was gently used. To complete the homogenisation and disrupt the embryos even further – solution was pushed through an insulin needle approximately two times. The suspension was filtered through 20 µm Cell Strainer. DAPI was added to reach 1 µg·ml⁻¹ working concentration and 25 000 G1 nuclei per sample were sorted into PBS (with protease inhibitor in low bind tubes) by flow-cytometer. The supernatant was removed after centrifugation (1000g, 15 min).

4.4.4.3 Flow-cytometry and nuclei sorting 24 DAP embryos and 4 DAG seedlings for ATAC-seq

Approximately ten 24 DAP embryos per sample were resuspended in LB01 buffer with 1x protease inhibitor and centrifuged (3000g, 15 min, 4 °C). Micro-blender was used for breaking down the embryos before filtering. DAPI was added to the suspension to reach 1µg·ml⁻¹ working concentration. After filtration of the suspension, 50 000 G1 nuclei per sample in PBS (with protease inhibitor in low bind tubes) were sorted by fluorescence-activated cell sorting (FACS). The supernatant was removed after centrifugation (1000g, 15 min).

4 DAG seedlings were prepared with one difference, the plant material was chopped with a razor blade on a glass petri dish.

4.4.4.4 ATAC-seq

Preparation of ATAC-seq libraries was done using ATAC-Seq Kit (Active Motif). The manufacturer's protocol was followed using 5-10 µl of tagmentase Tn5. Addition of Purification Binding Buffer preceded decross-linking with 1% SDS in TE. Amplification was done in 40 µl reactions, approximately 13 cycles per sample. Samples prepared using this protocol are ready for sequencing.

4.4.5 Sequencing

NovaSeq™ 6000 by Illumina was used to sequence both ChIP-seq and ATAC-seq samples. Sequencing was done on S4 flow cell with S4 Reagent Kit v1.5 (200 cycles).

4.4.6 Data analysis

For analysis, bash, a Unix shell was used (a command-line interface, a default shell for Linux operating systems).

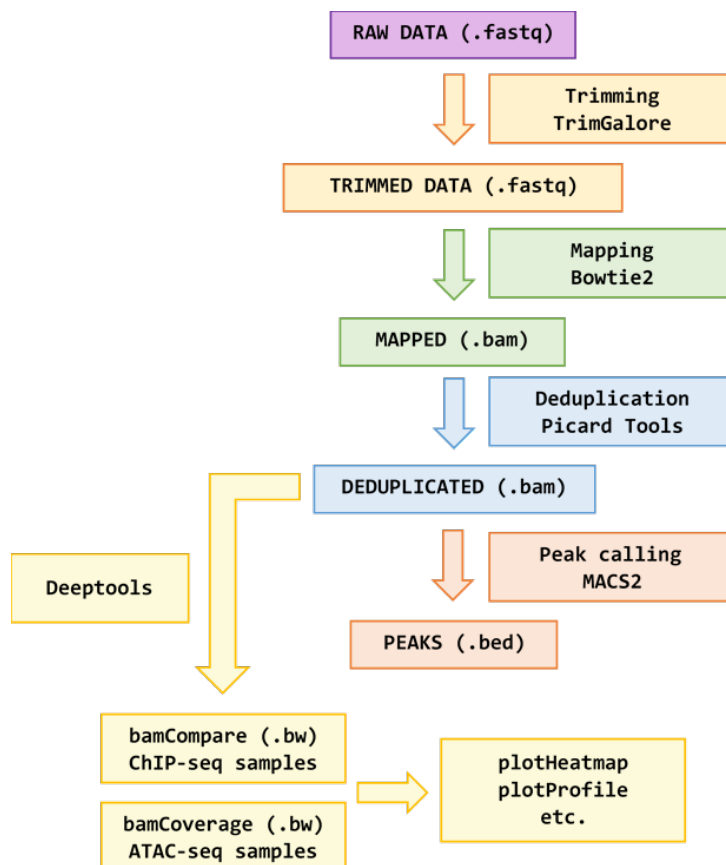


Figure 14: Data analysis workflow – individual steps taken (e.g. trimming), tools used for data adjustment (e.g. TrimGalore) and data file formats (e.g. .fastq).

4.4.6.1 ChIP-seq

Sequenced paired-end reads from separate lanes of the flowcell were concatenated for further data analysis.

```
cat mydata_R1_001.fastq.gz mydata_R1_001.fastq.gz > my_data_R1.fastq.gz
```

The following step was to check the quality of the sequenced reads with FastQC v0.11.5 and to check for sequencing adaptors and low-quality bases that may have affected mapping. Alternatively, multiQC was used to create a report for all samples analysed.

```
fastqc mydata_R1.fastq.gz  
multiqc .
```

Raw sequencing data were trimmed using TrimGalore v0.6.2. Trimming includes quality and adaptor trimming with integrated quality control using FastQC software.

```
trim_galore -paired -q 25 --fastqc mydata_R1.fastq.gz mydata_R2.fastq.gz -o  
trimmed -j 4
```

Bowtie2 v2.3.5.1 was used to index barley reference genome (bowtie2-build) which is necessary for mapping and alignment of reads onto reference sequences. Bowtie2 along with Samtools v1.10 was used to map and sort data.

```
bowtie2-build MorexV3_pseudomolecule.fasta MorexV3_pseudomolecule  
  
bowtie2 -p 12 --no-mixed --no-discordant -x MorexV3_pseudomolecule -1  
mydata_R1.fastq.gz -2 mydata_R2.fastq.gz 2>mapped/bowtie2_report.txt |  
samtools sort -o mapped/mydata_sorted.bam
```

Marking of duplicates by MarkDuplicates (Picard v2.9.0) is another crucial step. This tool locates and tags duplicate reads, these duplicate reads are defined as originating from a single fragment of DNA which arose during library construction using PCR.

```
java -jar picard.jar MarkDuplicates I=mydata_sorted.bam O=mydata_marked.bam  
M=marked_dup_metrics.txt REMOVE_DUPLICATES=true
```

Peak calling was performed by MACS2 v2.2.7.0 with normalization to input. Peak calling is used to identify regions in the genome that have been enriched with aligned reads as

a consequence of performing a ChIP-sequencing experiment.

```
macs2 callpeak -t mydata_marked.bam -c Input_marked.bam -f BAM -g 5.3e+9 --  
broad --outdir peaks
```

DeepTools v3.5.0 were used to further analyze data and plot profiles and heatmaps. Here, example command lines to generate 8DAP matrix and heatmap of ChIP-seq signal distribution around TSSes.

```
computeMatrix reference-point -S `cat 8DAP_bw_files.txt` -R  
broad_sharp_TSSes_CAGEall.bed -a 3000 -b 1000 -skipZeros -p 20 -o  
8DAP_ChIPseq_TSSes_mergedReps.mat.gz  
  
plotHeatmap --matrixFile 8DAP_ChIPseq_TSSes_mergedReps.mat.gz --outFileName  
8DAP_ChIPseq_TSSes_heatmaps.png --plotTitle "8DAP ChIP-seq profiles" --kmeans  
2
```

Bedtools v2.30.0 were used for peak manipulation (merge replicates, intersect samples and gene annotation).

```
bedtools intersect -a replica1.broadPeak -b replica2.broadPeak -sorted >  
intersect.broadPeak
```

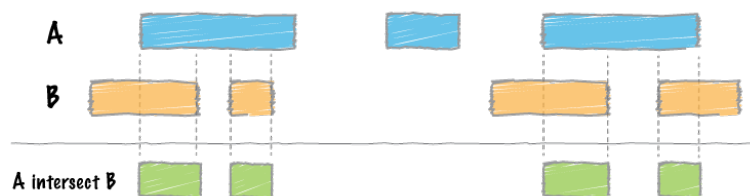


Figure 15: Visualisation of Bedtools option 'intersect'.

4.4.6.2 ATAC-seq

Data analysis steps including file concatenation, quality assessment, trimming, mapping to the barley reference and PCR duplicate removal were performed same as for the ChIP-seq data.

Samtools was used to sort and index data. The samtools fixmate tool corrects any flaws in read-pairing that may have been introduced by the aligner.


```
samtools sort -n -o mydata_marked_namesorted.bam mydata_marked.bam
samtools fixmate mydata_marked_namesorted.bam mydata_fixed.bam
```

Samtools and Bedtools v2.30.0 were used to convert .bam files into .bedpe format. Reads had to be shifted +4 bp and -5 bp for positive and negative strand respectively, to account for the 9 bp duplication created by DNA repair of the nick by Tn5 transposase and achieve base-pair resolution – bash script by Reske *et al.*, 2020, was used (bedpeTn5shift.sh). Similarly, for Minimal conversion (conversion of standard 10-column format BEDPE to the “minimal” format defined by MACS2), a bash script from Reske *et al.*, 2020, was used (bedpeMinimalConvert.sh).

```
samtools view -bf 0x2 mydata_fixed.bam | bedtools bamtoBED -i stdin -bedpe >
mydata_fixed.bedpe

chmod u+x bedpeTn5shift.sh
bash bedpeTn5shift.sh mydata_fixed.bedpe > mydata_tn5.bedpe

chmod u+x bedpeMinimalConvert.sh
bash bedpeMinimalConvert.sh mydata_tn5.bedpe > mydata_minimal.bedpe
```

Peak calling was performed by MACS2 v2.2.7.1. Bedtools v2.30.0 was used for peak manipulation (merge replicates, intersect samples and gene annotation). Bedtools intersect was used to find peaks present in both replicas (Figure 16).

```
macs2 callpeak -t mydata_minimal.bedpe -g 4.6e+9 --outdir peaks -n filename -
outdir peaks --broad -f BAMPE --broad-cutoff 0.05/0.01 --keep-dup all

bedtools intersect -a replica1.broadPeak -b replica2.broadPeak -sorted >
intersect.broadPeak
```

A DeepTools v3.5.0 tool computeMatrix was used to read coverages from ATAC-seq around TSS and arrange values overlapping the genomic features from the provided annotation file into a matrix. The annotation file “broad_sharp_TSSes_CAGEall.bed” is derived from CAGE data sets, which delineates the precise dominant TSSes for expressed genes (supplied by thesis supervisor Pavla Navrátilová, Ph.D). This matrix is further processed into a selected graphic (a heatmap here) by “plotHeatmap” tool.

```

computeMatrix    reference-point    -S    `cat    bw_files.txt`    -R
broad_sharp_TSSes_CAGEall.bed    --samplesLabel    8DAP_ATACseq    24DAP_ATACseq
4DAG_ATACseq    -a    3000    -b    1000    -skipZeros    -p    20    -o
All_ATACseq_mergedReps_genes.mat.gz

plotHeatmap    --matrixFile    All_ATACseq_mergedReps_genes.mat.gz    --outFileName
All_ATACseq_mergedReps_heatmaps.png    --plotTitle    "ATAC-seq profiles"    --yMax
0.25 0.3 0.8 --legendLocation none

```

For checking the correlation of sample replicates, we used multiBamsummary followed by plotCorrelation/plotPCA from DeepTools.

```

multiBamSummary    bins    -b    `cat    bam_filelist.txt`    --labels    8DAP_ATACseq_1
8DAP_ATACseq_2 24DAP_ATACseq_1 24DAP_ATACseq_2 4DAG_ATACseq_1 4DAG_ATACseq_2 -
--extendReads    --ignoreDuplicates    --maxFragmentLength    147    --centerRead    --
numberOfProcessors    30    -o    bam_correlation.npz

plotCorrelation    --corData    bam_correlation.npz    --corMethod    pearson    --whatToPlot
heatmap    --plotFile    ATACseq_bam_correlation.png    --skipZeros    --labels
8DAP_ATACseq_1 8DAP_ATACseq_2 24DAP_ATACseq_1 24DAP_ATACseq_2 4DAG_ATACseq_1
4DAG_ATACseq_2 --plotTitle    "ATACseq replica correlation"    --removeOutliers    --
plotHeight    10    --plotWidth    10    --plotNumbers

```

5 RESULTS

Barley embryos 8 DAP, 24 DAP and 4 DAG seedlings were collected and were subjected to ChIP-seq and ATAC-seq analysis. Sequencing data processing and analysis followed using bioinformatical tools.

For ATAC-seq a correlation plot was made using DeepTools. Specifically, the Pearson method was used to calculate correlation coefficients which can be seen in Figure 16. This correlation shows how similar replicas are to each other and how different samples and tissues vary. DeepTools principal component analysis (PCA) shows ATAC-seq replica correlation (the eigenvalues of the top two principal components; Figure 17).

With the help of DeepTools, coverage of ATAC-seq data around TSSes was visualised in a heatmap (Figure 18). High coverage on the left side of the TSS region usually signals for promoters. When comparing the three stages of barley studied, it is clear, that the highest coverage is present in 4 DAG seedlings; 8 DAP has the lowest coverage of the three.

The number of peaks for ATAC-seq for different tissues and replicas is summarized in Table 1. The intersection of peaks for replicas was done using Bedtools. For 4 DAG seedlings, bedtools merge was used since replica 2 showed significant variance from replica 1 (in other intersections, both outputs were taken into account). The highest number of ATAC-seq peaks can be seen in 24 DAP embryos.

All ChIP-seq samples correlated decently proving data reproducibility (Figure 19). Peak counting was used for histone mark ChIP-seq data sets (Table 2). Here it was distinguished between three types of tissues, and three types of histone modifications studied. Generally, a trend can be observed across different types of histone markers – the lowest number of peaks is found in 8 DAP embryos.

ChIP-seq histone mark enrichment around TSSes was visualised using DeepTools. This heatmap shows enrichment 1kb upstream and 3kb downstream of TSSes. Active, expressed genes (high enrichment of H3K9ac, H3K4me3), as well as repressed, inactive genes (high enrichment of H3K27me3) can be distinguished due to clustering (Figures 20, 21 and 22).

ATAC-seq data were integrated with histone marker ChIP-seq data. Typically, a positive correlation with active chromatin marks (H3K4me3, H3K9ac) and negative with inactive chromatin markers (H3K27me3) is observed. Coverage and enrichment for all types of tissues

for ATAC-seq and ChIP-seq data were observed with the help of the IGV browser. Along with these data sets, RNA-seq data sets were integrated as well (generated in collaborating research groups of M. Mascher and A. Pečinka). In the representative region of chromosome 3 (Chr3H:606.653.004-606.688.204) in Figure 23 different situations can be observed. HORVU.MOREX.r3.3HG0324810.1 gene is located in a repressed region of the genome – high H3K27me3 and low H3K4me3 with H3Kac, as well as low to none ATAC-seq peaks. HORVU.MOREX.r3.3HG0324830.1 gene on the other hand seems to be fully expressed as it shows high enrichment for H3K4me3, H3K8ac and ATAC-seq peaks for active chromatin along with peaks for the expressed region from RNA-seq data sets. Another region downstream of gene HORVU.MOREX.r3.3HG0324830.1 can be observed, which also shows signs of active chromatin and could possibly be an enhancer region since it does not show signs of expression.

The number of peaks for specific methods used are shown in Table 1 and 2. In Table 3 number of intersecting peaks for ATAC-seq and ChIP-seq data sets is presented - H3K4me3 and H3K9ac (both markers of active chromatin) intersection can be seen, as well as the intersection for ATAC-seq peaks, H3K4me3 and H3K9ac. In Table 3 the intersection for ATAC-seq peaks and genes +1000 bp upstream, along with no overlap for ATAC-seq and genes and promoters, is shown. From these peaks, the intersection for intergenic ATAC-seq peaks and H3K4me3 and H3K9ac were counted, which signify possible intergenic enhancer candidates – 506 for 8 DAP, 4 190 for 24 DAP and 3 617 for 4 DAG (Table 4).

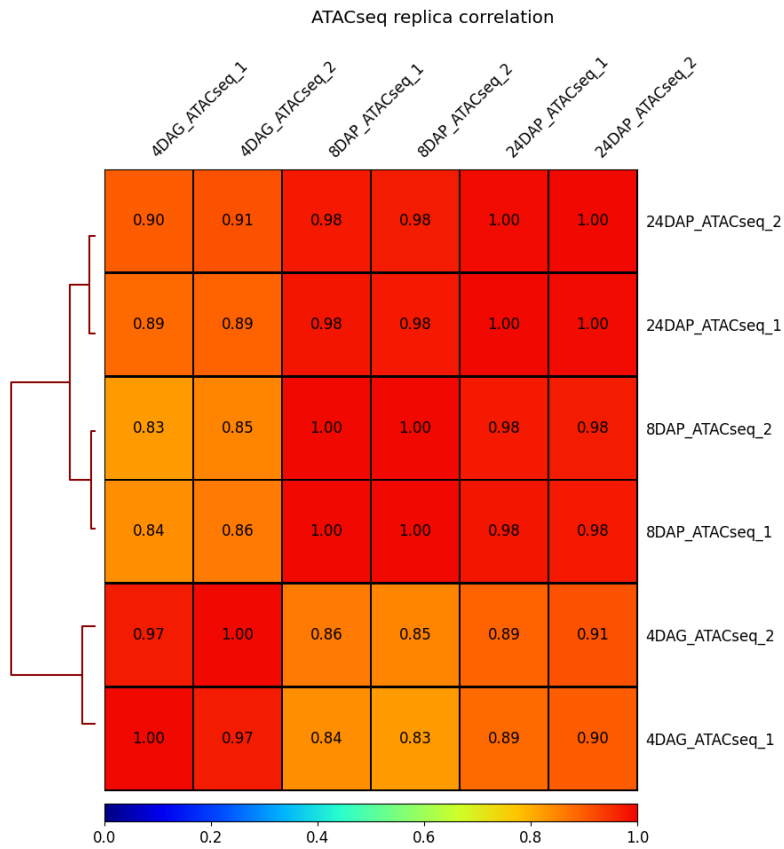


Figure 16: ATAC-seq data correlation heatmap, Pearson method. The correlation coefficient for samples describes the similarity between them (1.00 being the same sample, 0.0 having no correlation at all). This correlation most importantly describes the similarities and differences between replicas.

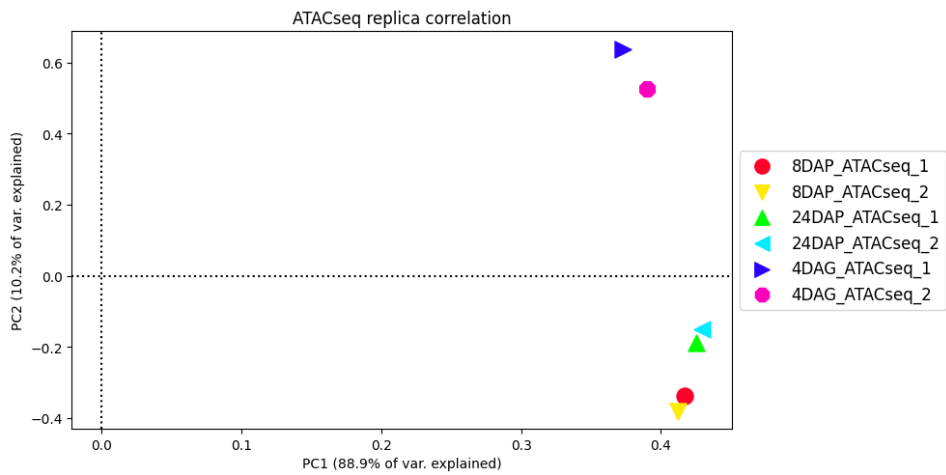


Figure 17: Principal component analysis (PCA) of ATAC-seq shows replica correlation with eigenvalues of the top two principal components.

ATAC-seq profiles

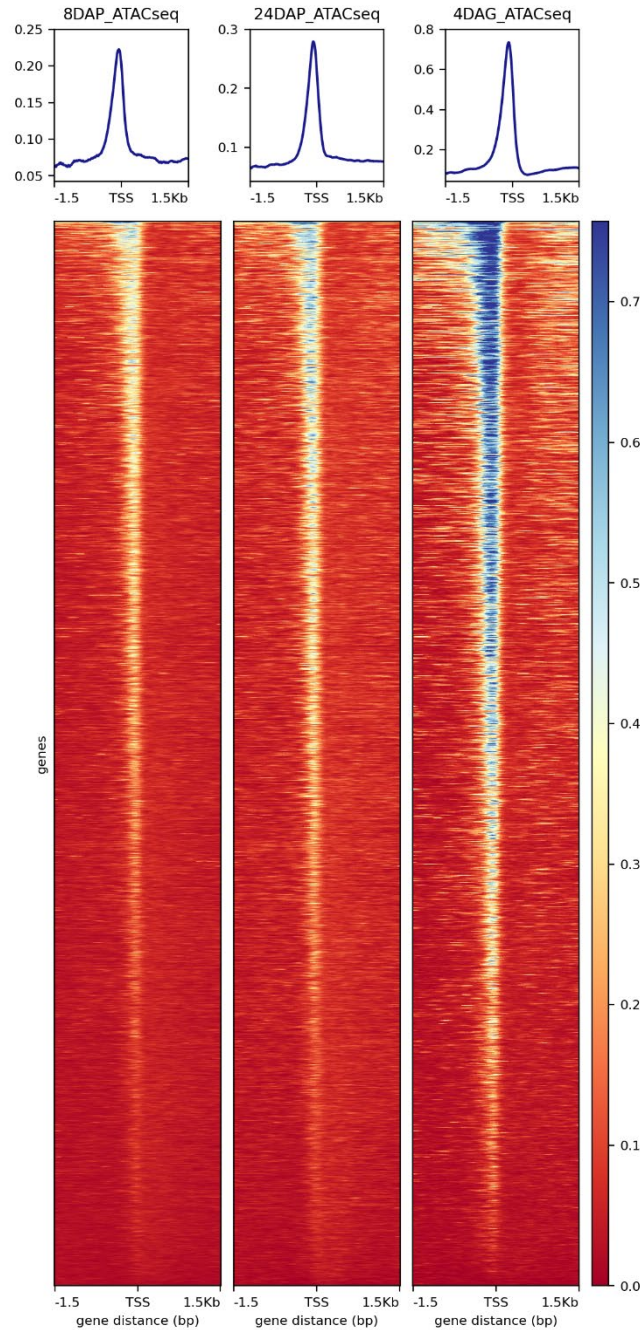


Figure 18: Heatmap of ATAC-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). All TSSes are aligned in the middle of the figure, 3 kb around the TSS are displayed. On top of the heatmap, a mean signal around the TSS is shown. Coverage is usually higher on the left side of the TSSes, which signals a promoter region.

Table 1: Number of ATAC-seq peaks for three different barley tissue types (8 DAP, 24 DAP, 4 DAG). Results of intersection (or pooling in case of 4 DAG) of the two replicates using Bedtools.

		Type of tissue		
ATAC-seq		8 DAP	24 DAP	4 DAG
Number of peaks	Replica 1	53 303	145 943	114 780
	Replica 2	55 758	127 166	29 863
	Intersect*	42 899	93 441	17 935
	Merge**	-	-	126 707

*Intersect = pairwise intersection between 2 data sets

**Merge = peaks from 2 data sets combined

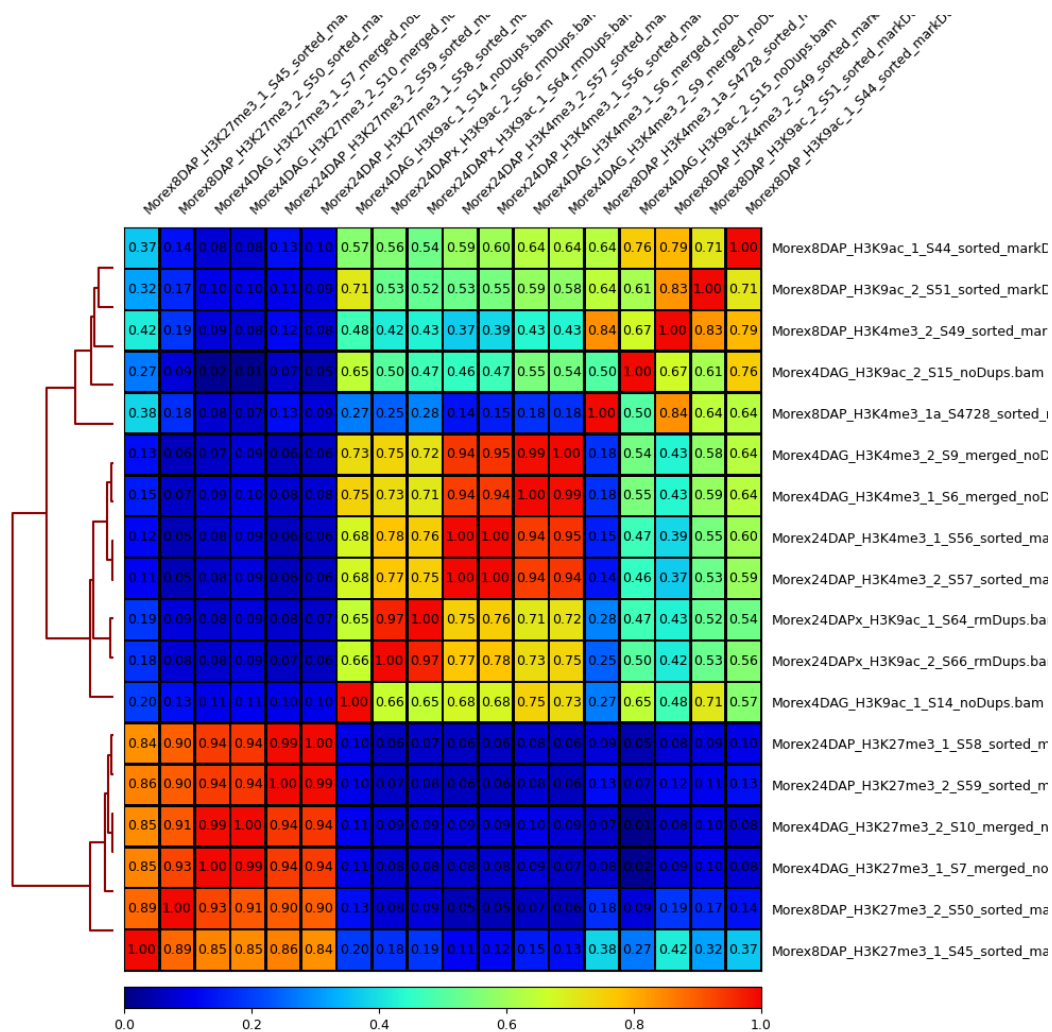


Figure 19: Correlation of mapped ChIP-seq reads, Pearson method. For ChIP-seq data – three tissue types (8 DAP, 24 DAP, 4 DAG), two replicates each, three histone modifications (H3K4me3, H3K9ac, H3K27me3). The correlation coefficient for samples describes the similarity between them (1.00 being the same sample, 0.0 having no correlation at all).

Table 2: Number of ChIP-seq peaks for three histone modifications (H3K9ac, H3K4me3, H3K27me3), three different barley tissue types (8 DAP, 24 DAP, 4 DAG). Two replicas each, the intersection of these replicas (Bedtools intersect).

		Type of tissue		
H3K9ac		8 DAP	24 DAP	4 DAG
Number of peaks	Replica 1	22 130	32 119	61 885
	Replica 2	26 393	66 104	30 178
	Intersect*	15 911	30 159	25 994
H3K4me3		8 DAP	24 DAP	4 DAG
Number of peaks	Replica 1	12 976	47 454	53 653
	Replica 2	16 198	51 563	48 396
	Intersect*	10 075	46 220	46 140
H3K27me3		8 DAP	24 DAP	4 DAG
Number of peaks	Replica 1	46 915	59 006	71 668
	Replica 2	43 595	51 373	85 421
	Intersect*	38 506	48 316	70 564

*Intersect = pairwise intersection between 2 data sets

8DAP ChIP-seq profiles

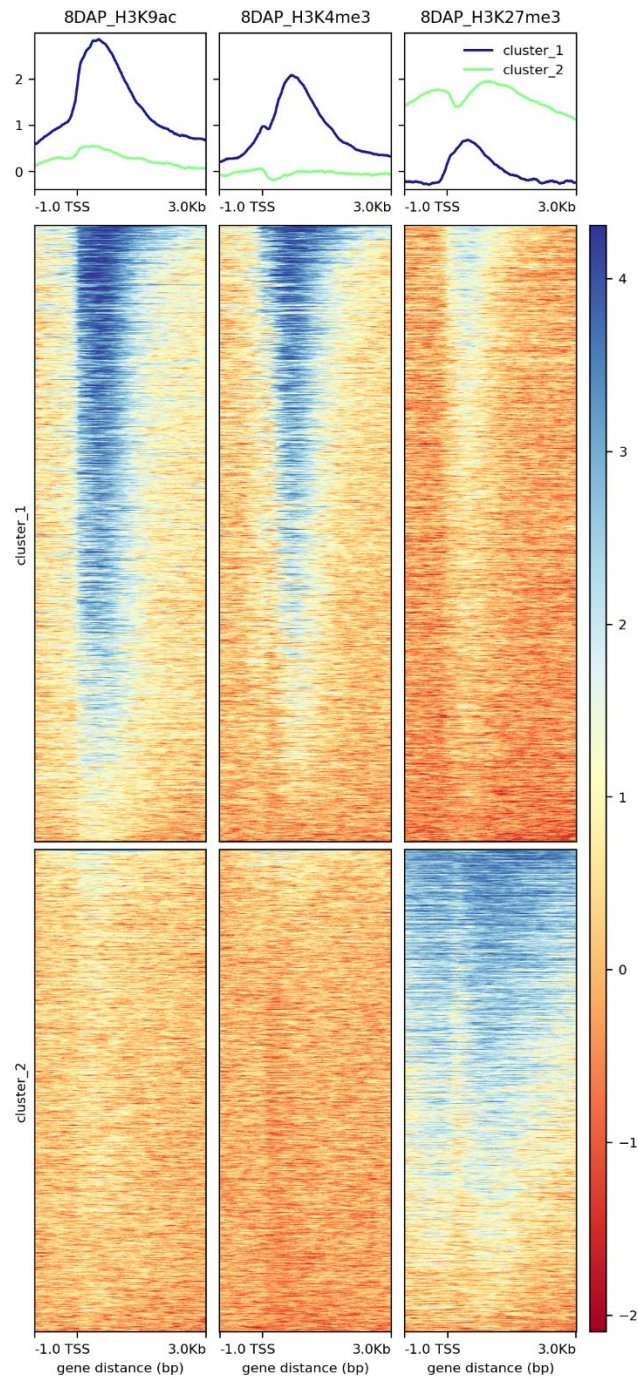


Figure 20: Heatmap of 8 DAP embryos ChIP-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 1 signifies expressed, active genes with high H3K4me3 and H3K9ac; cluster 2 shows unexpressed genes with high H3K27me3 enrichment.

24DAP ChIP-seq profiles

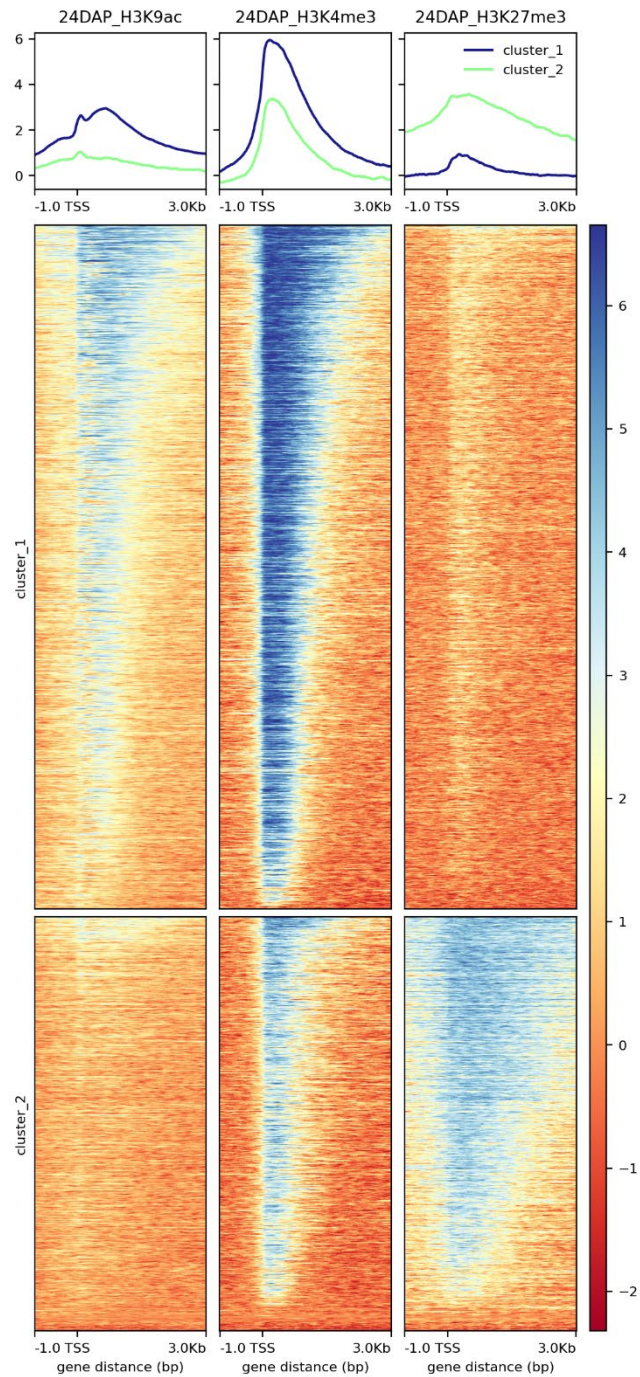


Figure 21: Heatmap of 24 DAP embryos ChIP-seq coverages around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 1 signifies expressed, active genes with high H3K4me3 and H3K9ac; cluster 2 shows unexpressed genes with high H3K27me3 enrichment and H3K4me3 enrichment.

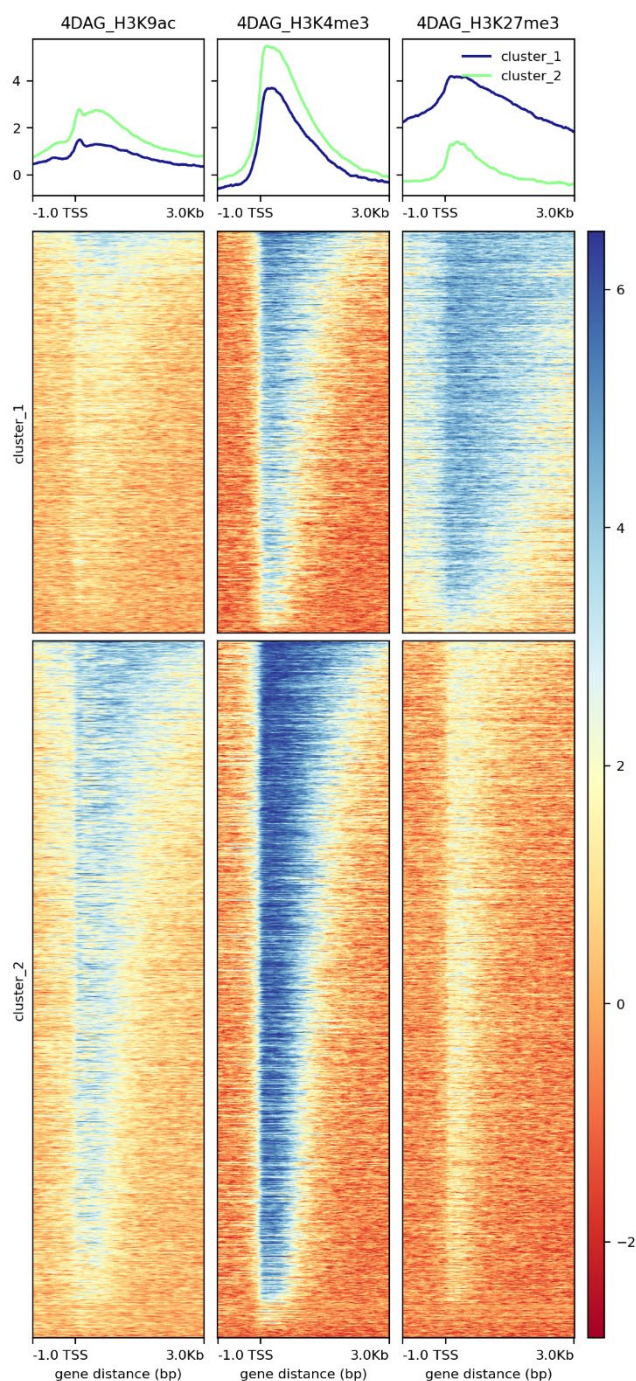


Figure 22: Heatmap of 4 DAG ChIP-seq around TSSes. Each line corresponds to a transcript. The coverage is summarized with a colour – red (no coverage) and blue (maximum coverage). Regions 1 kb upstream and 3 kb downstream around the TSSes are displayed. On top of the heatmap, a mean signal around the TSSes is shown. Two clusters are shown – cluster 2 shows unexpressed genes with high H3K27me3 enrichment and H3K4me3 enrichment; cluster 1 expressed, active genes with high H3K4me3 and H3K9ac.

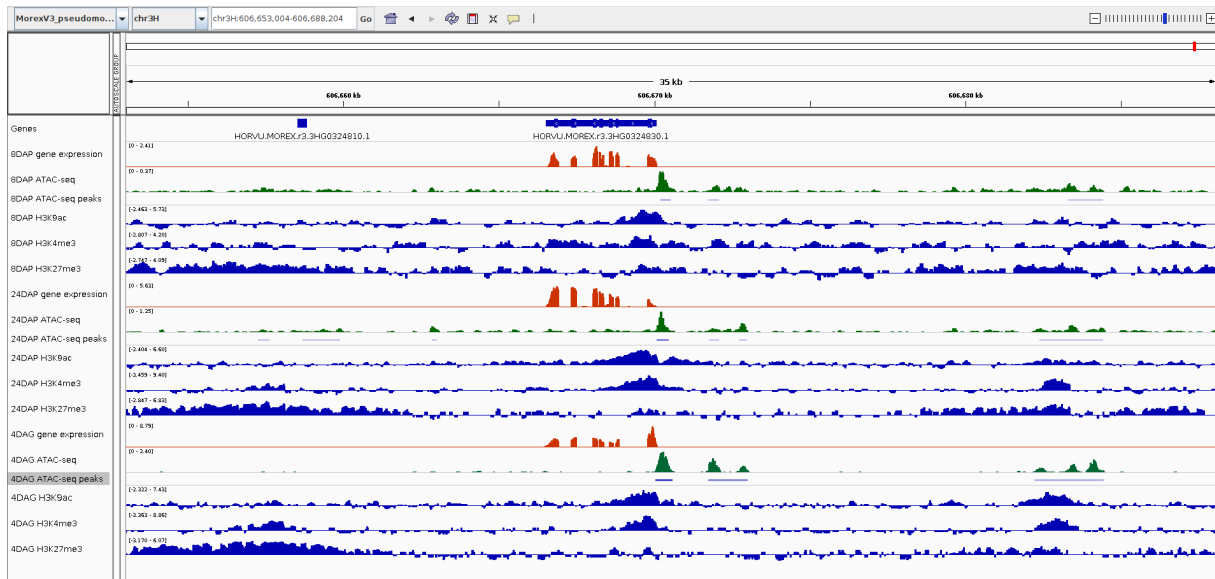


Figure 23: The IGV browser was used to display the epigenomic profiles across the whole barley genome. This specific region of chromosome 3 (chr3H:606 653 004-606 688 204) shows ChIP-seq coverage for H3K9ac, H3K27me3 and H3K4me3 and ATAC-seq peaks and coverage for three different types of tissues, 8 DAP, 24 DAP and 4 DAG. Along with ChIP-seq and ATAC-seq data sets, RNA-seq data sets were integrated, representing gene expression. HORVU.MOREX.r3.3HG0324810.1 gene is located in a repressed region of the genome – high H3K27me3 and low H3K4me3 with H3K9ac, as well as low to none ATAC-seq peaks. HORVU.MOREX.r3.3HG0324830.1 gene on the other hand seems to be fully expressed as it shows high enrichment for H3K4me3, H3K8ac and ATAC-seq peaks for active chromatin along with peaks for the expressed region from RNA-seq data sets. Another active chromatin region downstream of gene HORVU.MOREX.r3.3HG0324830.1 could possibly be an enhancer region since it does not show signs of expression.

Table 3: Data intersection for H3K4me3 and H3K9ac; ATAC-seq, H3K4me3 and H3K9ac; ATAC-seq with genes +1000bp upstream; ATAC-seq with no overlap with genes and promoters (intergenic).

		Type of tissue		
Data intersection		8 DAP	24 DAP	4 DAG
Number of peaks	Intersect H3K4me3/H3K9ac	8 878	27 447	23 577
	Intersect ATAC-seq/H3K4me3/H3K9ac	1 914	19 199	1 754
	Merge*			16 861
	Intersect ATAC-seq/genes +1000bp upstream	26 009	41 101	5 806
	Merge*			48 319
	ATAC-seq with no overlap with genes and promoters	21 045	56 656	12 825
	Merge*			85 174

*Intersection of ATAC-seq obtained by the merge option (see Table 1)

Table 4: Number of intergenic enhancer candidates for three types of barley tissue studied (8 DAP, 24 DAP, 4 DAG).

INTERGENIC ENHANCER CANDIDATES		Type of tissue		
		8 DAP	24 DAP	4 DAG
Number of peaks	Intergenic ATACseq/H3K4me3/H3K9ac	506	4 190	3 617

6 DISCUSSION

Genome-wide epigenetic profiling in plants has recently become one of the crucial types of experiments in molecular biology and plant science. A better understanding of epigenetics is important to facilitate novel and better approaches to crop improvement. The main objective of this thesis is to help decipher the epigenetic landscape of the barley genome, using ATAC-seq and ChIP-seq analysis. Three types of tissues (8 DAP, 24 DAP, 4 DAG) were used in order to compare expression patterns.

Through ATAC-seq, open chromatin was analysed across three different barley tissues. ATAC-seq provides a static assessment of chromatin architecture and reveals local, hyperaccessible regions of the genome. This method has proven to be a valuable tool in the identification of *cis*-regulatory elements in a variety of cell types (Shashikant *et E*ttensohn, 2019). Protocol for Active Motif kit was used, with one difference, embryos were fixed in formaldehyde since they were sorted by fluorescence-activated cell sorting (FACS). The pipeline for data analysis was followed (Reske *et al.*, 2020), with modifications.

Data analysis revealed that coverage of open chromatin around TSSes (Figure 18) is the highest in 4 DAG seedlings. The 4 DAG tissues are more abundant and it is easier to obtain good quality nuclei. The least amount of coverage is found in 8 DAP embryos, this might be due to the nature of the tissue itself. Alternatively, an argument can be made, that because of the challenging collection of sufficient amount of the material, the preparation and analysis of this tissue could yield less reliable data sets.

All samples were prepared in two replicas. How these replicas are similar to each other is shown via a correlation plot (Figure 16), which shows correlation coefficients, and principal component analysis (PCA) plot (Figures 17). Out of these, the least number of similarities is shown in 4 DAG tissues. This difference can be seen in Table 1 showing the number of ATAC-seq peaks in samples. One replica shows a significantly lower number of peaks and therefore a different bioinformatical tool (Bedtools ‘merge’) was used to work with these data sets. As the peaks called from the two replicates largely overlapped, the quality of both data sets is likely comparable. The difference is caused by the low sequencing depth in one of the replicates.

ChIP-seq analysis for H3K9ac, H3K4me3 and H3K27me3 aimed to locate enrichment of these histone post-translational modifications across the barley genome. These modifications help

distinguish active chromatin from inactive one. The marks of active chromatin tend to be H3K9ac and H3K4me3, usually present at TSSes, in expressed genes, but also at enhancer regions (Duan *et al.*, 2018; Jiang *et al.*, 2020). Repressed chromatin mark H3K27me3 works generally in the opposite manner (Duan *et al.*, 2018). Data analysis followed a similar process to ATAC-seq data, however, the Tn5 shift in reads was omitted. A correlation plot for replicas was made (Figure 19). In summary, these replicas are very similar in nature and allow further analysis. The number of peaks for various tissues and histone modifications were counted, and 8 DAP embryos are yet again a type of tissue with the lowest enrichment for all marks (Table 2).

ATAC-seq and ChIP-seq analyses in and of themselves are not enough to conduct a proper genome-wide profiling analysis. Other methods and data sets need inclusion to interpret data correctly (Yan *et al.*, 2020). In this thesis, RNA-seq data sets were often added to generate profiles and interpret data sets for expressed regions (these data sets were generated in collaborating research groups of M. Mascher and A. Pečinka). Their importance lies in the unreliable nature of high-fidelity reference sequence, which does not often contain all genes. RNA-seq data sets help locate genes not present in the reference sequence and therefore aid in distinguishing expressed sequences with active chromatin marks from enhancer sites. Among data not included in this thesis, but no less important, are for instance Hi-C (Kong *et al.*, 2019) or DNA methylation analysis (Liu *et al.*, 2020). However, interpretation of these data sets is beyond the scope of this thesis.

Based on the data included in this thesis, many active enhancer candidates were selected. These candidates had to meet specific conditions to be considered for future analysis. High H3K4me3 and H3K9ac enrichment, low H4K27me3 enrichment and ATAC-seq peaks indicative of active chromatin were the main requirements (Figure 23). Other conditions included no overlap with genes and promoters (Table 3). It must be taken into account, that intronic enhancers exist (Borsari *et al.*, 2021), but peak analysis would be too complicated to conduct and therefore does not fall within the scope of this thesis. These established conditions were met by 506 regions in 8 DAP tissues, 4 190 in 24 DAP tissues and 3 617 in 4 DAG (Table 4).

Further studies of these enhancer candidates are needed and their identification is the first step in decoding the epigenetic landscape of barley. Along with analysing other types of datasets, target genes for these enhancers must be found to confirm their activity status. The final proof of enhancer functionality would desirably come from using stable or transient reporter expression assays *in vivo* (Lin *et al.*, 2019).

7 CONCLUSION

Analysis of the barley genome was performed through ATAC-seq and ChIP-seq methods followed by analysis using bioinformatical tools. Specific conditions were established to distinguish between promoter and enhancer regions. Both of these regions are characterised by high H3K4me3 and H3K9ac enrichment, low H4K27me3 enrichment, ATAC-seq peaks for transposase-accessible chromatin (active chromatin). However, no overlap with genes and promoters was a condition met only by enhancer regions.

Based on these conditions 506 regions in 8 DAP tissues, 4 190 in 24 DAP tissues and 3 617 in 4 DAG tissues could be considered in future studies as potential enhancer regions. This selection is but a first step for further investigation. Analysis of other data sets, locating enhancer target genes and proofs of enhancer functionality must be performed to prove the validity of these findings.

8 REFERENCES

- Andersson R., Sandelin A. (2020): Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics* 21: 71–87.
- Baile F., Gómez-Zambrano Á., and Calonje M. (2022): Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. *Plant Communications* 3:100267.
- Borsari B., Villegas-Mirón P., Pérez-Lluch S., Turpin I., Laayouni H., Segarra-Casas A., Bertranpetit J., Guigó R., Acosta S. (2021): Enhancers with tissue-specific activity are enriched in intronic regions. *Genome Research* 31(8):1325-1336.
- Charron J. B., He H., Elling A. A., Deng X. W. (2009): Dynamic landscapes of four histone modifications during de etiolation in *Arabidopsis*. *Plant Cell* 21:3732–3748.
- Chen X., Bhadauria V., Ma B. (2018): ChIP-Seq: A Powerful Tool for Studying Protein-DNA Interactions in Plants. *Current Issues in Molecular Biology* 27:171–180.
- Cortijo S., Charoensawan V., Roudier F., Wigge P. A. (2018): Chromatin Immunoprecipitation Sequencing (ChIP-Seq) for Transcription Factors and Chromatin Factors in *Arabidopsis thaliana* Roots: From Material Collection to Data Analysis. *Methods in Molecular Biology* 1761:231–248.
- Duan C. G., Zhu J. K., Cao X. (2018): Retrospective and perspective of plant epigenetics in China. *Journal of Genetics and Genomics*. 45: 621–638.
- Furey, T. (2012): ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics* 13: 840–852.
- Gehring M. (2013): Genomic imprinting: insights from plants. *Annual Review of Genetics* 47:187–208.
- Grossniklaus U., Paro R. (2014): Transcriptional silencing by Polycomb-group proteins. *Cold Spring Harbor Perspectives in Biology* 6(11):a019331.
- He Y. (2009): Control of the transition to flowering by chromatin modifications. *Molecular Plant* 2:554–564.
- Huang X., Pan Q., Lin Y., Gu T, Li Y. (2020): A native chromatin immunoprecipitation (ChIP)

- protocol for studying histone modifications in strawberry fruits. *Plant Methods* 16(10).
- Huang Y., Sicar S., Ramirez-Prado J. S., Manza-Mianza D., Antunez-Sanchez J., Brik-Chaouche R., Rodriguez-Granados N. Y., An J., Bergounioux C., Mahfouz M. M., Hirt H., Crespi M., Concia L., Barneche F., Amiard S., Probst A. V., Gutierrez-Marcos J., Ariel F., Raynaud C., Latrasse D., Benhamed M. (2021): Polycomb-dependent differential chromatin compartmentalization determines gene coregulation in *Arabidopsis*. *Genome Research* 31(7):1230–44.
- International Barley Genome Sequencing Consortium, Mayer K. F., Waugh R., Brown J. W., Schulman A., Langridge P., Platzer M., Fincher G. B., Muehlbauer G. J., Sato K., Close T. J., Wise R. P., Stein N. (2012): A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426): 711–6.
- Iwafuchi-Doi M., Zaret K. S. (2014): Pioneer transcription factors in cell reprogramming. *Genes & Development* 28(24): 2679–2692.
- Jenuwein, T., Allis, C. D. (2001): Translating the histone code. *Science* 293: 1074–1080.
- Jiang D., Kong N. C., Gu X., Li Z., He Y. (2011): *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLOS Genetics* 7(3):e1001330.
- Jiang J., Ding A. B., Liu F., Zhong X. (2020): Linking signalling pathways to histone acetylation dynamics in plants. *Journal of Experimental Botany* 71(17): 5179–5190.
- Kapazoglou A., Tondelli A., Papaefthimiou D., Ampatzidou H., Francia E., Stanca M. A., Bladenopoulos K., Tsaftaris A. S. (2010): Epigenetic chromatin modifiers in barley: IV. The study of barley Polycomb group (PcG) genes during seed development and in response to external ABA. *BMC Plant Biology* 10(73): 1471–2229.
- Kong S., Zhang Y. (2019): Deciphering Hi-C: from 3D genome to function. *Cell Biology and Toxicology* 35(1):15–32
- Kolovos P., Knoch T. A., Grosveld F. G., Cook P. R., Papantonis A. (2012): Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin* 5:1.
- Kumar S., Bucher, P. (2016): Predicting transcription factor site occupancy using DNA

- sequence intrinsic and cell-type specific chromatin features. *BMC Bioinformatics* 17, S4.
- Kusch S., Panstruga R. (2017): Mlo-based resistance: an apparently universal “weapon” to defeat powdery mildew disease. *Molecular Plant-Microbe Interactions* 30: 179–189.
- Langridge P. (2018): Economic and Academic Importance of Barley. In: Stein N., Muehlbauer G. (eds.) *The Barley Genome. Compendium of Plant Genomes*. Springer, Cham (2018).
- Li S., Lin D., Zhang Y., Deng M., Chen Y., Lv B., Li B., Lei Y., Wang Y., Zhao L., Liang Y., Liu J., Chen K., Liu Z., Xiao J., Qiu J. L., Gao C. (2022): Genome-edited powdery mildew resistance in wheat without growth penalties. *Nature* 602(7897): 455–460.
- Lin Y., Meng F., Fang C., Zhu B., Jiang J. (2019): Rapid validation of transcriptional enhancers using agrobacterium-mediated transient assay. *Plant Methods* 15:21.
- Liu C., Lu F., Cui X., Cao, X. (2010): Histone methylation in higher plants. *Annual review of plant biology* 61: 395–420.
- Liu P., Slotkin R. K. (2020): Cis-regulatory units of grass genomes identified by their DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* 117(41): 25198–25199.
- Liu X., Yang S., Yu C. W., Chen C. Y., Wu K. (2016): Histone Acetylation and Plant Development. *Enzymes* 40: 173–199.
- Luger K., Mader A. W., Richmond R. K., Sargent D. F., and Richmond T. J. (1997): Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648): 251–260.
- Mascher M., Wicker T., Jenkins J., Plott C., Lux T., Koh C. S., Ens J., Gundlach H., Boston L. B., Tulpová Z., Holden S., Hernández-Pinzón I., Scholz U., Mayer K. F. X., Spannagl M., Pozniak C. J., Sharpe A. G., Šimková H., Moscou M. J., Grimwood J., Schmutz J., Stein N. (2021): Long-read sequence assembly: a technical evaluation in barley. *The Plant Cell* 33(6):1888–1906.
- Morell P. L., Clegg M. T. (2007): Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the United States of America* 104: 3289–3294.

- Navrátilová P., Toegelová H., Tulpová Z., Kuo Y.-T., Stein N., Doležel J., Houben A., Šimková H., Mascher M. (2022): Prospects of telomere-to-telomere assembly in barley: analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnology Journal*. doi: 10.1111/pbi.13816. Epub ahead of print. PMID: 35338551.
- Newton A. C., Flavell A. J., George T. S., Leat P., Mullholland B., Ramsay L., Revoredo-Giha C., Russell J., Steffenson B. J., Swanston J. S., Thomas W. T. B., Waugh R., White P. J., Bingham I. J. (2011): Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Security* 3(141).
- Nowicka A., Kovacik M, Tokarz B., Vrána J., Zhang Y., Weigt D., Doležel J., Pečinka A. (2020): Dynamics of endoreduplication in developing barley seeds. *Journal of Experimental Botany* 72(2): 268–282.
- Nowicka A., Sahu P. P., Kovacik M., Weigt D., Tokarz B., Krugman T., Pecinka A. (2021): Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel. *Genes* 12(5):711.
- Ogbourne S., Antalis T. M. (1998): Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochemical Journal* 331: 1–14.
- Olsen O.-A. (2001): Endosperm development: cellularization and cell fate specification. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 233–267.
- Paro R., Grossniklaus U., Santoro R., Wutz A. (2021): Introduction to Epigenetics. *Learning Materials in Biosciences*, Springer.
- Pourkheirandish M., Hensel G., Kilian B., Senthil N., Chen G., Sameri M., Azhaguvel P., Sakuma S., Dhanagond S., Sharma R., Mascher M., Himmelbach A., Gottwald S., Nair S. K., Tagiri A., Yukuhiro F., Nagamura Y., Kanamori H., Matsumoto T., Willcox G., Middleton C. P., Wicker T., Walther A., Waugh R., Fincher G. B., Stein N., Kumlehn J., Sato K., Komatsuda T. (2015): Evolution of the grain dispersal system in barley. *Cell* 162:527–539.
- Reske J. J., Wilson M. R., Chandler R. L. (2020): ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics & Chromatin* 13(22).

- Rodríguez M. V., Barrero J. M., Corbineau F., Gubler F., Benech-Arnold R. L. (2015): Dormancy in cereals (not too much, not so little): about the mechanisms behind this trait. *Seed Science Research* 25: 99–119.
- Rotasperti L., Sansoni F., Mizzotti C., Tadini L., Pesaresi P. (2020): Barley's second spring as a model organism for chloroplast research. *Plants* 9(7): 803.
- Sabelli P. A., Larkins B. A. (2009): The development of endosperm in grasses. *Plant Physiology* 149: 14–26.
- Sato K. (2020): History and future perspectives of barley genomics. *DNA Research* 00(0): 1–8.
- Shashikant T., Etensohn C. A. (2019): Genome-wide analysis of chromatin accessibility using ATAC-seq. *Methods of Cellular Biology* 151:219–235.
- Shlyueva D., Stampfel G., Stark A. (2014): Transcriptional enhancers: from properties to genome-wide predictions. *Natural Reviews Genetics* 15(4): 272–86.
- Schulte D., Close T. J., Graner A., Langridge P., Matsumoto T., Muehlbauer G., Sato K., Schulman A. H., Waugh R., Wise R. P., Stein N. (2008): The international barley sequencing consortium (IBSC) – at the threshold of efficient access to the barley genome. *Plant Physiology* 149: 142–147.
- Sreenivasulu N., Borisjuk L., Junker B. H., Mock H.-P., Rolletschek H., Seiffert U., Weschke W., Wobus U. (2010): Barley grain development: toward an integrative view. In: Jeon K, (ed.) *International review of cell and molecular biology*. Elsevier Inc., Volume 281: 49–89.
- Tan H., Liu T., Zhou T. (2022): Exploring the role of eRNA in regulating gene expression. *Mathematical Biosciences and Engineering* 19(2):2095–2119.
- Tan M., Luo H., Lee S., Jin F., Yang J. S., Montellier E., Buchou T., Cheng Z., Rousseaux S., Rajagopal N., Lu Z., Ye Z., Zhu Q., Wysocka J., Ye Y., Khochbin S., Ren B., Zhao Y. (2011): Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. 146(6):1016–28.
- Varshney R. K., Pandey M. K., Chrikineni A. (2018): *Plant Genetics and Molecular Biology. Advances in Biochemical Engineering/Biotechnology*, Springer, Switzerland.

- Wang Y. P., Cheng X., Shan Q. W., Zhang Y., Liu J. X., Gao C. X., Qiu J. L. (2014): Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* 32: 947–951.
- Weber H., Borisjuk L., Wobus U. (2005): Molecular physiology of legume seed development. *Annual Review of Plant Biology* 56: 253–279.
- Weber B., Zicola J., Oka R., Stam M. (2016): Plant Enhancers: A Call for Discovery. *Trends in Plant Science* 21(11): 974–987.
- Xiao T., Li X., Felsenfeld G. (2021): The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proceedings of the National Academy of Sciences of the United States of America* 118(7):e2023127118.
- Yan F., Powell D. R., Curtis D. J., Wong N. C. (2020): From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. *Genome Biology* 21(22).
- Zhao G., Wang J., Chen X., Sha H., Liu X., Han Y., Qiu G., Zhang F., Fang J. (2022): OsASHL1 and OsASHL2, two members of the COMPASS-like complex, control floral transition and plant development in rice. *Journal of Genetics and Genomics* S1673-8527(22)00079-0.
- Zhao T., Zhan Z., Jiang D. (2019): Histone modifications and their regulatory roles in plant development and environmental memory. *Journal of Genetics and Genomics*. 46(10): 467–476.