Ghent University – Faculty of Sciences

Department of Plant Biotechnology and Bioinformatics

VIB Center for Plant Systems Biology

Development of AP-MS tools to explore the chromatin landscape and low affinity interactions in plants.

Caroline Matthijs

Promoters: Dr. Jelle Van Leene and Prof. Dr. Geert De Jaeger

Thesis submitted as partial fulfillment of the requirements to obtain the degree of Doctor of Philosophy in Sciences: Biochemistry and Biotechnology.

Academic year 2019-2020

GHENT UNIVERSITY

FACULTY OF SCIENCES

VIB-UGENT CENTER FOR PLANT SYSTEMS BIOLOGY

Examination board

Promoters

Prof. Dr. Geert De Jaeger
Ghent University - Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant System Biology

Dr. Jelle Van Leene
Ghent University - Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant System Biology

Chair

Prof. Dr. Daniël Van Damme
Ghent University - Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant System Biology

Secretary

Dr. Laurens Pauwels
Ghent University - Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant System Biology

Other voting members

Dr. Jefri Heyman
Ghent University - Department of Plant Biotechnology and Bioinformatics
VIB - Department of Plant System Biology

Prof. Dr. Sven Eyckerman
Ghent University – Department of Biomolecular Medicine
VIB – Center for Medical Biotechnology
Cancer Research Institute Ghent

Prof. Dr. Wolfgang Dröge-Laser
Julius-Maximilians-Universität Würzburg - Department of Pharmaceutical Biology

Acknowledgments

"Success is not final, failure is not fatal.
It is the courage to continue that counts."

W. Churchill

Six years ago I decided to start a PhD at the research group of Prof. Geert De Jaeger, exploring the possibilities to generate an *in planta* gene-centered protein-DNA analysis tool. Although Geert warned me from the beginning, little did I know how challenging and complex this project would be. Luckily, I had the most amazing colleagues and it is because of a fantastic team work I can present this doctoral thesis today.

From the start it has been a bumpy road with a lot of expectations but equally disappointments. However, the support and positive feedback of Geert and Jelle after every single meeting boosted me to continue the search for (lets be honest) 'the holy grail'. Especially the little pats on the back and the small pep talks of Geert at moments of frustration and renouncement, gave me enough courage to pursue. Thank you Geert for letting me discover a strong personal quality, my perseverance.

I could not have imagined a better mentor than Jelle. I'm amazed of the enormous amount of knowledge he possesses and I appreciate his critical mindset. Although he has a buzzy agenda, working on different projects, guiding master students and mentoring other PhD's, I could always knock on his 'sidewall' for a question or second opinion. I also could always count on the other fantastic group members, Astrid, Nancy, Eveline, Geert, Bernard, Dominique, Nienke, Michiel and Lana. For every practical (and even non-work related) issue, help was offered. They supported me when workload was high, when the disappointments became too much, when I became pregnant,... Especially a big thank you for Nancy and Astrid, whom put a lot of energy in validating my final results and are to date still performing experiments for this doctoral research. The fact that I can present this thesis, is because of you guys. Without your fantastic support and friendship, I would not have reached the finish of this bumpy road.

Daarnaast wil ik ook mijn ouders bedanken. Alhoewel het voor jullie nog steeds niet 100% duidelijk is wat ik de afgelopen 6 jaar heb gedaan, bleven jullie al die jaren interesse en bezorgdheid tonen. Bedankt voor de vele warme maaltijden en gezellige babbels aan de keukentafel.

Als laatste wil ik natuurlijk mijn man Brecht bedankten. De afgelopen jaren waren heftig. We verhuisden drie keer, bouwden een huis samen, we gaven elkaar het ja-woord en mochten anderhalf jaar geleden onze prachtige dochter Louise verwelkomen. Dit combineren met een doctoraat was niet altijd even evident. Bedankt voor de vele knuffels, jouw inzet en oprechte interesse. Je bent een fantastische man en een super-papa-held!

Summary

To gain a deeper insight on how an organism functions, the complex interplay between macromolecules needs to be mapped. A precise regulation of the genetic information encoded in DNA by proteins is of utmost importance for the proper functioning of an organism. An appropriate transcription, which can be both time and stimuli dependent, results in cellular functions that allow for changes in growth, differentiation, division and responses to environmental impulses. Furthermore, proteins are essential laborers of a biological system and interact with each other, forming complexes and initiating signaling cascades for proper biological activity. Different methods exist to unravel the identity and location of protein-DNA interactions (PDIs, **chapter 1**) or to map binary and high-throughput protein-protein interactions (PPIs, **chapter 2**). While the list of PPI and PDI analysis tools keeps expanding, there still remains unexplored territory. For plant PDIs a gene-centered method is not available yet, making the analysis of a specific plant genomic region dependent on ectopic Y1H screens or *in vitro* methods. On the other hand, studying transient PPIs by applying affinity purification is challenging as most of the time only the stable components are retrieved and the weak, transient interactions are lost.

In this doctoral research we tackled both shortcomings by testing different strategies in analogy with previously reported methods in mammalian cells and yeast. We examined three different strategies to develop an *in planta* gene-centered PDI tool. First, we tested a multifunctional T-DNA construct making use of a tagged exogenous DNA binding protein and its corresponding DNA binding element for the specific isolation of different promoter sequences and its interacting partners (**chapter 3**). We identified the *in vivo* formaldehyde crosslinking step as the culprit for the pull-down of a massif amount of non-specific interactions, making it impossible to retrieve bona-fide interactions with mass spectrometry. To avoid crosslinking, we combined the gene-centered strategy with proximity labeling, allowing *in vivo* biotinylation and a stringent streptavidin-based purification of interacting proteins (**chapter 3**). Although an initial experiment did not result in a desirable outcome, further well-considered adjustments specific for the analysis of interactions in the dens nuclear organelle can lead to an optimal approach. As a third option, we tested the applicability of CRISPR for the endogenous targeting and subsequent isolation of multicopy loci and their proteome (**chapter 4**). Also with this approach we encountered difficulties, however we propose that implementing the proximity-based labeling strategy can result in a preferred gene-centered method for the identification of PDIs in plants.

Additionally, we explored different options to reveal weak transient PPIs in plants. We validated an in-house developed pull-down technique (AP-MS), identifying the weak transient interactions between the anaphase promoting complex (APC) and its mitotic substrates (**chapter 5**). Our data strengthens the hypothesis for a mitotic function of the APC co-activator CCS52B and revealed known and new putative APC substrates. In addition, we tested if proximity labeling could further expand the CCS52B interactome and validate the AP-MS results.

In conclusion, in this doctoral research we validated different strategies to tackle interactomic challenges related to plants. Our data reveals that no single approach is superior and is accompanied with minor or major drawbacks. We provide new insights in the development of an *in planta* gene-centered method and report on the functionality of a pull-down protocol for elucidating weak transient interactions. We also laid a foundation for further research on the mitotic role of the APC co-activators.

Samenvatting

Om een beter inzicht te krijgen hoe een organisme functioneert, is het nodig om de complexe samenwerking tussen macromoleculen in kaart te brengen. Een preciese regulatie van de genetische informatie door eiwitten is van groot belang voor het correct functioneren van een organisme. Een correcte transcriptie, die tijds- en stimuli-afhankelijk kan zijn, resulteert in cellulaire processen die veranderingen in groei, differentatie, celdeling en reacties op omgevingsfactoren toelaten. Bovendien zijn eiwitten essentiële arbeiders van een biologisch systeem en interageren ze met elkaar om macromoleculaire complexen te vormen en signaalcascades te initiëren voor een correct functioneren. Er bestaan verschillende methoden om eiwit-DNA interacties (PDI's, **hoofdstuk 1**), of binaire en high-throughput eiwit-eiwit interacties (PPI's, **hoofdstuk 2**) te identificeren en te localiseren. Hoewel de lijsten met PDI en PPI analysetechnieken blijven uitbreiden, blijft er nog onontgonnen terrein. Er is geen gen-gecentreerde methode voorhanden om PDI's te analyseren in planten, waardoor het onderzoeken van een specifieke genomische regio van planten afhankelijk is van yeast one-hybrid analyses of *in vitro* toepassingen. Anderzijds is het identificeren van tijdelijke en zwakke PPI's door het toepassen van affiniteitszuivering een uitdaging, aangezien vaak enkel sterke interacties behouden blijven tijdens de zuivering en de zwakke, tijdelijke interacties verloren gaan.

In dit doctoraatswerk hebben we beide tekortkomingen aangepakt door verschillende strategieën uit te testen, in analogie met eerder gepubliceerde methoden in dierlijke cellen en gist. We bestudeerden drie verschillende strategieën voor de ontwikkeling van een plantspecifiek gen-gecentreerde PDI techniek. Eerst hebben we het gebruik van een multifunctioneel T-DNA construct getest waarbij we een getagd exogeen DNA-bindend eiwit en het overeenkomstige DNA-bindingselement aanwenden voor de specifieke isolatie van verschillende promoter sequenties en hun interagerende partners (**hoofdstuk 3**). We identificeerden de *in vivo* formaldehyde crosslinking stap als de boosdoener voor het weerhouden van een gigantische hoeveelheid niet-specifieke interacties waardoor het achterhalen van bonafide interacties met massaspectrometrie onmogelijk werd. Om crosslinking te vermijden, hebben we deze gen-gecentreerde strategie gecombineerd met proximity labeling, waardoor interagerende eiwitten *in vivo* gebiotinyleerd worden en opgezuiverd worden op basis van de sterke affiniteit tussen biotine en streptavidine (**hoofdstuk 3**). Ondanks een initieel experiment niet tot het gewenste resultaat heeft geleid, kunnen toekomstige weloverwogen aanpassingen, specifiek voor de analyse van interacties in het dense nucleaire organel, leiden tot een optimale aanpak. Als een derde optie hebben we de

toepasbaarheid van CRISPR uitgetest voor de endogene targeting en daaropvolgende isolatie van multicopy loci en hun proteoom (**hoofdstuk 4**). Ook bij deze toepassing ondervonden we moeilijkheden, maar we postuleren dat in combinatie met proximity labeling deze strategie kan resulteren in een krachtige gen-gecentreerde methode voor de identificatie van PDI’s in planten.

Daarnaast hebben we verschillende opties onderzocht om zwakke, tijdelijke PPI's in planten te onthullen. We hebben een in eigen huis ontwikkelde techniek voor affiniteitszuivering (AP-MS) gevalideerd, door zwakke, tijdelijke interacties te identificeren tussen het anaphase promoting complex (APC) en zijn mitotische substraten (**hoofdstuk 5**). Onze data versterkt de hypothese voor een mitotische functie van de APC co-activator CCS52B en onthult gekende en nieuwe mogelijke APC substraten. Daarnaast zijn we ook nagegaan of proximity labeling het CCS52B interactoom verder kon uitbreiden en of het de AP-MS resultaten kon valideren.

Concluderend, in dit doctoraatswerk hebben we verschillende strategieën gevalideerd om interactomic uitdagingen met betrekking tot planten aan te pakken. Onze resultaten tonen aan dat geen enkele toepassing superieur is en dat het gepaard gaat met kleine of grote nadelen. We bieden nieuwe inzichten aan in de ontwikkeling van een plantspecifiek gen-gecentreerde methode en rapporteren over de functionaliteit van een pull-down protocol voor het ophelderen van zwakke, tijdelijke interacties. We hebben ook een basis gelegd voor verder onderzoek naar de mitotische rol van de APC co-activators.

Table of contents

List of abbreviations

AD	activator domain
AP	affinity purification
APC	anaphase promoting complex
APEX	ascorbate peroxidase
ARF	auxin response factors
ATP	adenosine triphosphate
Aux/IAA	auxin/indole3-acetic acid proteins
AuxRE	auxin response elements
BE	binding element
BiFC	bimolecular fluorescence complementation
Bp	base pair
BRET	bioluminescence resonance energy transfer
CAK	CDK-activating kinase
CAPTURE	CRISPR affinity purification in situ of regulatory elements
CBP	calmodulin binding protein
CCS52	cell cycle switch 52
CDC20	cell division cycle 20
CDH1	CDC20 homolog 1
CDK	cyclin dependent kinase
cDNA	complementary DNA
CFP	cyan fluorescent protein
ChAP	chromatin affinity purification
ChIP	chromatin Immunoprecipitation
Co-IP	co-immunoprecipitation
CRE	cis-regulatory elements
CRISPR	clustered regularly interspaced short palindromic repeats
Cub	C-terminal part of ubiquitin
CY3	cyanine 3
DBD	DNA binding domains
DBP	DNA binding proteins
dCas9	death/deactivated Cas9
DDR	DNA damage response
DNA	deoxyribonucleic acid
DPI	DNA-protein Interaction
DSB	double stranded break
DUB	deubiquitylating enzyme
dUTP	deoxyuridine triphosphate
E1	ubiquitin-activating enzyme
E2	ubiquitin-conjugating enzyme
E3	ubiquitin-protein ligase
EMSA	electro mobility shift assay
enChIP	engineered DNA binding molecule-mediated chromatin immunoprecipitation
ETS	external transcribed spacer
FACT	facilitates chromatin transcription
FDR	false discovery rate
FRET	fluorescence resonance energy transfer

GENECAPP	Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics
GFP	green fluorescent protein
GloPro	genomic locus proteomics
GRN	gene regulatory networks
gRNA	guide RNA
GUS	β-glucuronidase
HA	human influenza hemagglutinin
HAD	histone deacetylases
hCas9	human codon optimized Cas9
iChIP	insertional chromatin immunoprecipitation
IgG	immunoglobulin G
IP	immunoprecipitation
JA	jasmonate
Kd	dissociation constant
Km^R	kanamycin resistance
LacI	lac repressor
LacO	lactose operator
LC-MS/MS	liquid chromatography-tandem MS
LNA	locked nucleic acid
MBD	microtubule-binding domain
mKG2	kusabira-green fluorescent mutant
MMQPCR	monochrome multiplex quantitative PCR method
MS	mass spectrometry
NAA	synthetic auxin
NEB	nuclear envelope breakdown
NGS	next-generation sequencing
NLS	nucleus localization sequence
Nub	N-terminal part of ubiquitin
ORF	open reading frame
P2H	protoplast two-hybrid
PAM	protospacer-adjacent motif
PBM	protein binding microarray
PCR	polymerase chain reaction
PDI	protein-DNA interactions
PIC	pre-initiation complex
PICh	proteomics of isolated chromatin segments
POL I	DNA-dependent RNA polymerase I
POL II	DNA-dependent RNA polymerase II
POL III	DNA-dependent RNA polymerase I
PPB	preprophase band
PPI	protein-protein interaction
ProtA	protein A
ProtG	protein G tag
PSB-D	*Arabidopsis* cell suspension cultures
PTM	posttranslational modifications
qPCR	quantitative polymerase chain reaction
rDNA	ribosomal DNA
Rhino	rhinovirus 3C protease cleavage sites
RNA	ribonucleic acid

rRNA	ribosomal RNA
RTA	repressed transactivator
RT-PCR	reverse transcription polymerase chain reaction
SBP	streptavidin-binding peptide
SELEX	systematic evolution of ligands by exponential enrichment
SICAP	selective isolation of chromatin-associated proteins
SILAC	stable isotope labeling by amino acids in cell culture
TAL	transcription activator-like
TAP	tandem affinity purification
TAS	telomere-Associated Sequence
TBP	TATA-box-binding protein
TChAP	tandem chromatin affinity purification technique
T-DNA	transfer DNA
TEV	tobacco etch virus
TF	transcription factor
TOC	translocon complex
TOR	target of rapamycin
Ub	ubiquitin
UBP	ubiquitin-specific protease
UV	ultraviolet
Y1H	yeast 1-hybrid
Y2H	yeast two-hybrid
Y3H	yeast three-hybrid
YFP	yellow fluorescent protein

Chapter 1

Exploring the chromatin proteome.

Caroline Matthijs[1,2], Astrid Gadeyne[1,2], Jelle Van Leene[1,2], Geert De Jaeger[1,2]

Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

Author contributions

CM performed literature study and wrote the chapter. AG, JVL and GDJ supervised and complemented the writing.

Abstract

Transcription regulation is the most important biological function of a cell and is primarily controlled by TFs and other regulatory proteins who interact with proximal and distal regulatory regions on the genome. Unraveling the identity and location of these transcription regulators is of utmost importance to understand an organism. Different tools exist to analyze protein-DNA interactions (PDIs), either focusing on the DNA binding protein or exploring the protein interactome of a specific genomic locus. In this chapter an overview of these different techniques is given along with advantages, challenges and some successful applications in different species.

Introduction

A precise regulation of the genetic information encoded in DNA is of utmost importance for the proper functioning of an organism. An appropriate transcription, which can be time and stimuli dependent, results in cellular functions that allow for changes in growth, differentiation, division and responses to environmental impulses. The basis of this important regulation is the sequence-specific binding of proteins on DNA. These protein-DNA interactions (PDIs) are essential and mutations in one of the two interaction partners can lead to severe malfunctioning of the organism. Unraveling these PDIs is therefore of great value to understand gene regulation.

Transcription is regulated by the interplay of different DNA binding proteins (DBPs) (Carlberg and Molnár, 2014). Transcription factors (TFs) directly bind DNA at specific sequence motifs present in promoter regions or in distal cis-regulatory elements (CREs) like enhancers. When bound, a TF will initiate a cascade of DBP recruitments influencing transcription in two ways. One cascade leads to the rearrangement of the chromatin, the other results in transcription activation. TF-based recruitment of co-activators (or co-repressors), like chromatin remodeling complexes and histone modifiers, helps to rearrange chromatin. Histone modifiers are responsible for the post transcriptional modifications on the N-terminal tail of histones. Depending on the modification and the position on the tail, transcription is activated or repressed (Hsieh and Fischer, 2005; Li et al., 2007). These modifications form a histone code that is recognized by chromatin remodeling complexes. A well-known plant chromatin remodeling complex is the SWI/SNF complex (Pfluger and Wagner, 2007). This complex weakens histone-DNA interactions via ATP hydrolysis, resulting in the shift or displacement of nucleosomes. This can lead to the formation of euchromatin, a less compact form allowing accessibility of cis-regulatory elements, or the formation of heterochromatin, a compact form associated with inactive genes. Secondly, TFs recruit the general initiation factors and DNA-dependent RNA polymerase II (POL II) to the regulatory TATA-box element leading to the formation of the pre-initiation complex (PIC) and the start of transcription (Nikolov and Burley, 1997). Another important protein complex involved in transcription regulation is the mediator complex. This co-activator complex forms a bridge between TFs that bind distal cis-regulatory elements and the promoter region with the associated PIC (Szutorisz et al., 2005). Figure 1 displays an overview of the different components needed for transcriptional regulation. It is clear that the chromatin proteome is a complex interplay of different proteins and complexes resulting in the specific regulation of the genetic code.

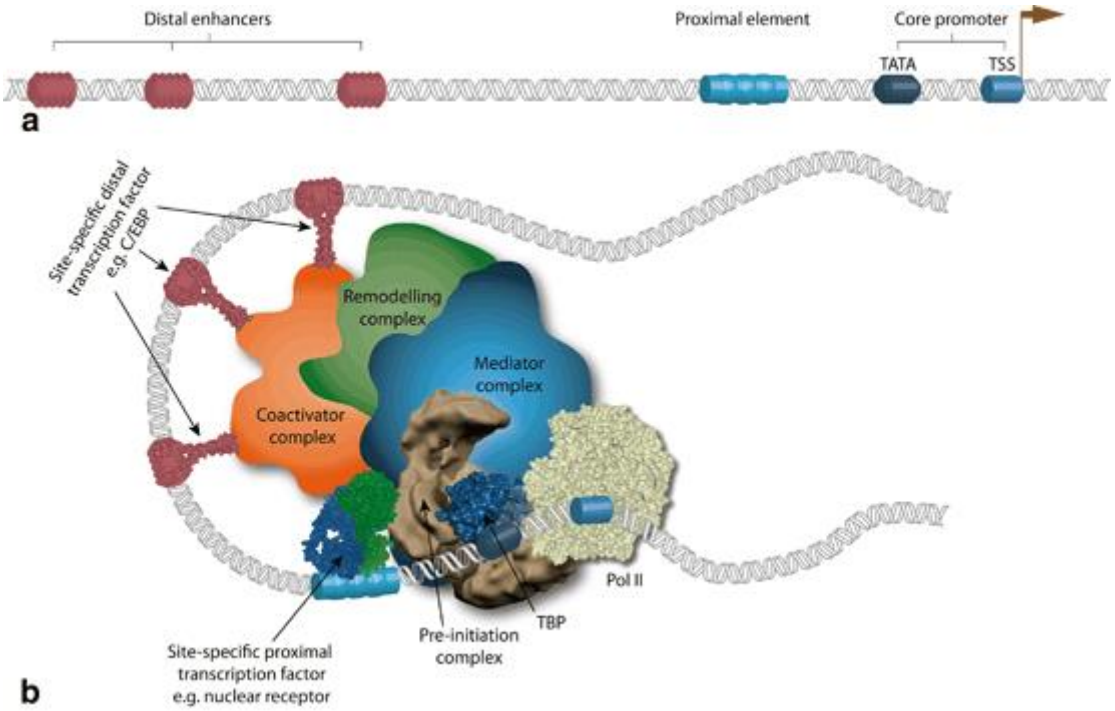


Figure 1: Components of the transcriptional regulation. **a)** Position of different DNA binding elements essential for transcriptional regulation. TSS: transcription start site. **b)** The different proteins and complexes that are needed for initiation of transcription. Pol II: RNA polymerase II. TBP: TATA-box binding protein. Figure adopted from Carlberg and Molnár, 2014.

TFs are the key regulators of transcription. Identifying them and their specific target genes is crucial to understand the genetic regulation of an organism. From the year 2000 on, several groups attempted to identify the collection of TFs of many plant species (Riechmann et al., 2000; Riano-Pachon et al., 2007; Richardt et al., 2007; Mochida et al., 2009; Romeuf et al., 2010; Mochida et al., 2013; Jin et al., 2017) which are all documented in the publicly available plant TF database PlantTFDB (http://planttfdb.cbi.pku.edu.cn/). Based on protein sequences, DNA binding domains (DBD) and Chromatin Immunoprecipitation (ChIP)-seq data, TFs have been identified and classified in different families. At this moment, 320,370 TFs from 165 plant species have been identified and classified in 58 TF families (Jin et al., 2017). For *Arabidopsis* 1,770 transcription factors have been identified and grouped in 50 different families (Yilmaz et al., 2011). However, the associated consensus binding sites for many of these TFs remain largely unknown and as a consequence the function and target genes stay uncharacterized.

Identifying protein-DNA interactions could help to establish a link between TFs and cis regulatory elements. The techniques used for the identification of PDIs can be divided in two groups, the transcription factor centered and the gene-centered ones. TF-centered techniques identify the DNA loci that are bound by the TF proteins. Gene-centered techniques on the other hand identify the proteins binding to a specific DNA locus. In this chapter an overview of these different techniques is given along with advantages, challenges and some successful applications in different species.

TF-centered tools

TF footprinting

DNase I footprinting is a quick and easy way to determine TF occupancy and identification of cis-regulatory elements (motifs) in the genome. Active regulatory regions bound by TF's have a more open DNA structure which makes it accessible for DNA nucleases like DNase I. These hypersensitive regions will be cut by the DNase I in to small pieces, which can be sequenced and aligned to the genome to determine the genomic footprint. Within these hypersensitive regions some sites will have a lower sequencing level, indicating the binding of a TF. TF binding will prevent the DNase I from cutting, leaving marks in the hypersensitive region that represent the binding motifs (figure 2). If one has a possible TF candidate for the identified binding motif, DNase I footprinting can again be applied to confirm this assumption. To do this, the candidate TF is added to a synthetic DNA sequence representing the hypersensitive region. Adding DNase I to this mixture will lead to the cutting of the DNA sequence with exception of the TF-DNA interaction location. This will result in a mixture of different DNA sequence lengths that can be visualized on a DNA gel. The area where there is an interaction will result in a blind spot on the gel. DNA without the potential binding TF is used as a control. Performing this DNase I footprinting under different conditions can also shed light about the specific TF-DNA interaction requirements.

A genomic DNase I footprinting experiment has been applied for rice (Zhang et al., 2012), where they observed that most hypersensitive regions were located in the area of a promoter and that some hypersensitive regions were tissue specific. Similar results were obtained in *Arabidopsis thaliana* (Zhang et al., 2012). Combining the genomic *Arabidopsis* DNase I hypersensitive regions with ChIP-seq data revealed a nice overlap indicating that these hypersensitive regions are putative regulatory elements.

An alternative footprinting strategy has been developed, called assay for transposase-accessible chromatin using sequencing (ATAC-seq), which not only elucidates TF binding sites but also the occupancy of nucleosomes (Buenrostro et al., 2013). ATAC-seq uses a prokaryotic transposase, Tn5, loaded with adapters for high-throughput DNA sequencing, to fragment and label active genomic regions. Subsequent amplification and NGS allows identification of these active regions, TF binding sites and nucleosome positions. Combining this transposase strategy with NGS results in a faster and cheaper footprinting protocol using a limited amount of cell material compared to DNase I footprinting. Recently a protocol for ATAC-seq on isolated plant

nuclei has been reported and has been used for mapping accessible chromatin regions in different plant species (Bajic et al., 2018; Maher et al., 2018).

TF footprinting is an easy and straightforward approach to validate TF-DNA interactions but it has been shown that TFs with a short DNA residence time have no footprints (Sung et al., 2014). Therefore, TF footprinting is a good option to determine the genomic footprint, identifying active cis-regulatory DNA and visualizing the binding of stable TF-DNA interactions. But additional techniques are needed to determine the complete transcription regulatory network of an organism.

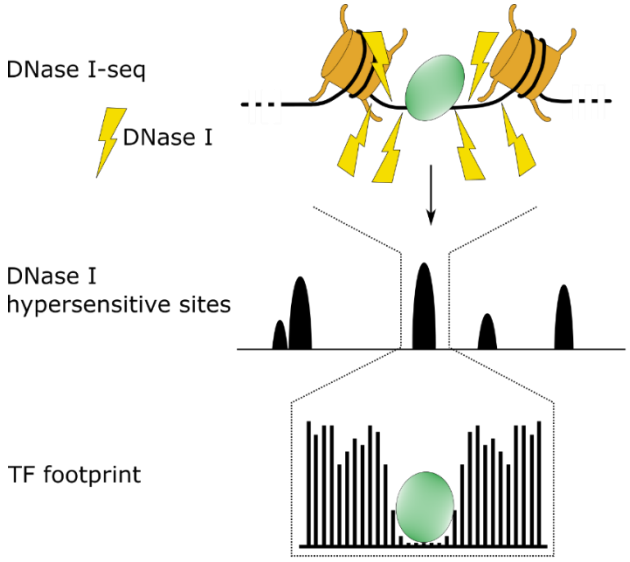


Figure 2: Schematic description of DNase I footprinting. Hypersensitive regions will be cut by the DNase I in to small pieces, which can be sequenced and aligned to the genome to determine the genomic footprint. Within these hypersensitive regions some sites will have a lower sequencing level, indicating the binding of a TF

Systematic Evolution of Ligands by Exponential Enrichment

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Tuerk and Gold, 1990; Gijs et al., 2016) entails an *in vitro* screening technique where a group of random 30 to 80-mer DNA-sequences is tested for binding with a specific protein. Iterative rounds of binding, separation and amplification are performed to isolate the DNA sequences with high affinity and specificity. A library of up to 10^{16} unique sequences is first incubated with the protein of interest, followed by separation of the bound sequences from the unbound sequences. This critical step can be achieved by immobilizing the protein on magnetic beads or by fusing it to an affinity tag. Also EMSA (electro mobility shift assay) can be applied for the isolation of bound sequences. The next step is the elution of the bound sequences from the protein and amplification via a polymerase chain reaction (PCR) to generate an enriched pool of sequences for the next selection round. Typically, 10 to 15 rounds of selection are performed to isolate

high affinity sequences. These DNA sequences are then identified via next generation sequencing.

SELEX is a useful *in vitro* technique when working with highly purified proteins that bind strongly and with a high specificity to their DNA-binding motif. This requirement cannot always be fulfilled by the protein of interest. Also the obtained DNA sequences are rather a prediction since SELEX is performed outside the native context of the protein. An *in vivo* validation, like Chromatin Immunoprecipitation (ChIP) is therefore recommended.

SELEX is not only used for identification of protein-DNA interactions. Small molecules, viruses, whole cells and tissues can be used for DNA sequence selection. Also random RNA sequences or fractionated genomic DNA can be used. Figure 3 displays an overview of all possible modifications to SELEX.

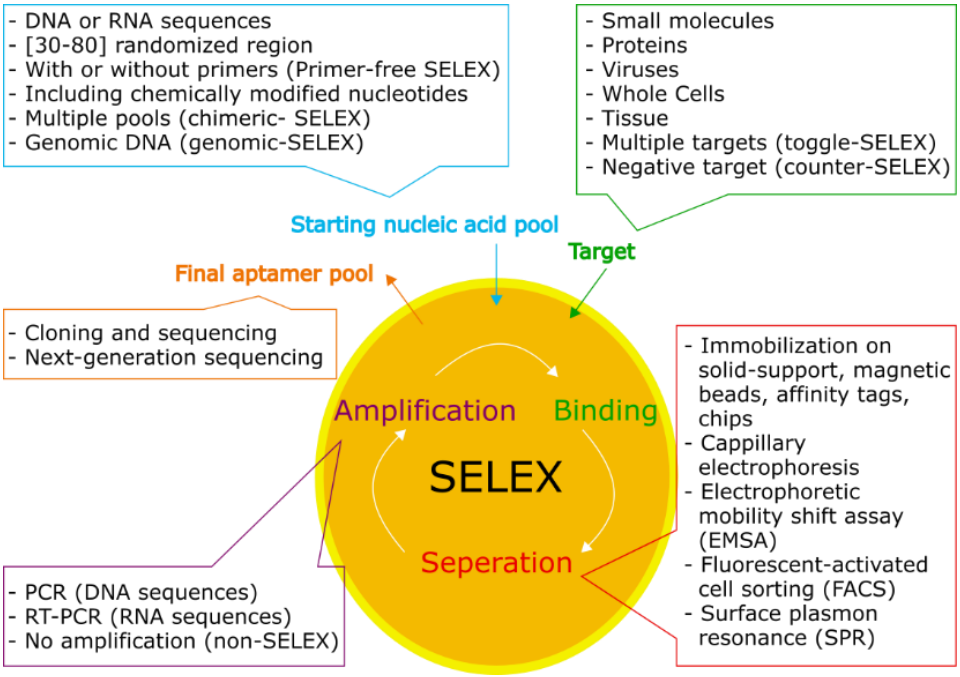


Figure 3: Overview of the most often used variations on the basic SELEX protocol to select new binding motifs. Consecutive rounds of binding, separation and amplification results in the identification of a final pool of aptamers binding the target. Different targets and nucleic acid pools can be applied, as well as different strategies to perform separation and amplification. Figure adapted from Gijs et al., 2016.

SELEX has been used for identification of the DNA binding sites of different plant TF's. 45 DNA binding motifs identified via SELEX can be found in the open source JASPAR database. Liang et al. used SELEX to identify the consensus binding motif of the plant SBP Domain Transcription Factor AtSPL14 (Liang et al., 2008). He isolated 20 dsDNA sequences capable of binding AtSPL14, all containing a core sequence of "CCGTAC". Also the binding sites of the *Arabidopsis* class I TCP proteins; which are developmental

regulators that influences the growth of leaves, stems and petioles, and pollen development; was specified by SELEX (Viola et al., 2011). Even more, they could show that all of the class I TCP proteins bind the binding site of the type GTGGGNCCNN, except for TCP11 that showed a preference for the sequence GTGGGCCNNN. All interactions were validated by performing an electrophoretic mobility shift assay (EMSA) and Yeast 1-hybrid (Y1H) assays.

Protein binding microarray

When a big amount of DNA binding proteins need to be analyzed, SELEX becomes time-consuming. An alternative exists, called protein binding microarray (PBM) analysis (Andrilenas et al., 2015). Recombinant TF's labeled with a tag are loaded on a microarray which contains random synthetic double-stranded DNA fragments. Binding of the TF on its corresponding binding motif can be visualized by a fluorescent labeled anti-tag antibody, producing the final spot intensity image. Aligning the corresponding dsDNA sequences will result in the identification of the binding motif. Because the density of dsDNA sequences can vary per spot and microarray, a control experiment is necessary. Here a second fluorescent signal is detected, originated from a dsDNA-specific fluorescent dye like Sybrgreen I or Cyanine 3 (CY3) linked to dUTP (Wang et al., 2011). These signals of density are then subtracted from the TF binding signals, resulting in normalized signal intensities (figure 4).

One drawback of PBM is the limited amount of binding motif that can be present on one chip. To circumvent this, microarrays are developed containing dsDNA fragments build up by overlapping binding motifs. PBM11 is such an example containing at least four times more sequences than previous PBMs (Godoy et al., 2011). PBM11 contains all possible 11-mer double-stranded oligonucleotide combinations that could be recognized by any TF, as well as all lower order k-mers (i.e. 6-mers, 7-mers, etc.). Because of its large number of dsDNA sequences, PBM11 is highly fit for high-throughput analysis of any eukaryotic TF. Godoy used PBM11 to identify the binding sites of MYC2 (G-box) and ERF1 (GCC-box). Not only the highest-affinity binding sites were identified, also variants with lower affinity were detected. Many more PBM analysis were executed for plant TFs and the results are brought together in an open source database called UniPROBE (Hume et al., 2015).

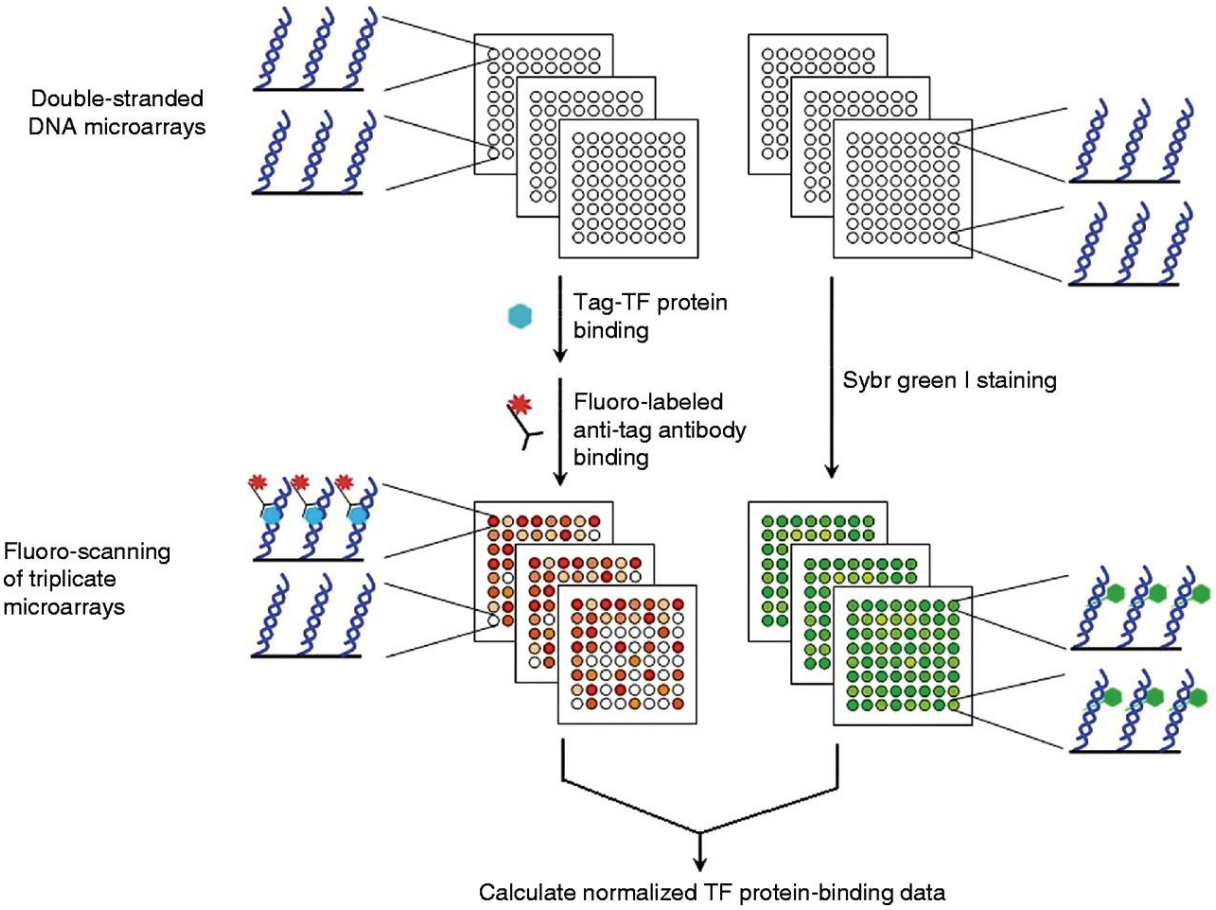


Figure 4: Schematic description of PBM. In a first step high-density dsDNA microarray's are produced. Secondly, a pool of purified TF protein of interest tagged with an epitope is allowed to bind directly to the microarray's. After removing unbound proteins, a fluorophore-conjugated antibody specific to the epitope is added to bind the TF's residing on the microarray. To confirm the reproducibility of dsDNA microarray detection, the experiments are performed in triplicate. To eliminate the influence of the density of dsDNA probes a second fluorescent signal is detected, originated from the dsDNA-specific fluorescent dye Sybrgreen I. These signals of density are then subtracted from the TF binding signals, resulting in normalized signal intensities. Figure adopted from Wang et al. 2011

DNA-Protein Interaction ELISA

An alternative for PBM is DNA-Protein Interaction ELISA (DPI-ELISA), specifically developed for automation (Brand et al., 2013). Here a 96 or 384 well microtiter plate with streptavidin-coated wells is used instead of a microarray. Biotinylated dsDNA fragments are individually immobilized in the wells after which HIS-epitope tagged DNA binding proteins are added. Binding is visualized via photometric detection, similar to PBM. As validation of this new screening strategy, they confirmed the interaction between the well-characterized Arabidopsis WRKY11 DNA-binding protein with the high affinity W-box TTGACY binding motif. They also were able to identify the consensus DNA binding element for the uncharacterized DBP, AtTIFY1.

DNA affinity purification sequencing

Another high-throughput TF-DNA-binding assay is DNA affinity purification sequencing (DAP-seq) (O'Malley et al., 2016; Bartlett et al., 2017). This technique uses *in vitro* expressed affinity-tagged TFs to capture genomic DNA sequences, followed by next generation sequencing (NGS) allowing for the generation of a genome-wide binding-site map. In a first step, genomic DNA is fragmented and ligated to sequencing adaptors, generating a genomic DNA library. In parallel, a TF of interest fused to an affinity tag is expressed in vitro and isolated using the appropriate affinity resin. Subsequently, the immobilized TF is combined with the genomic DNA library allowing sequence-specific binding. After removal of unbound DNA sequences, bound DNA fragments are released, amplified, sequenced and mapped on the genome. One DNA library can be applied for several DAP-seq experiments, and in combination with the high throughput *in vitro* expression of a library of TF open reading frames (ORFs), high throughput DAP-seq analyses can rapidly generate a genome wide overview of TF binding sites, also called a cistrome. DAP-seq analysis of 529 *Arabidopsis* TFs resulted in a cistrome which highly matches with previously *in vivo* identified TF binding sites (O'Malley et al., 2016). Another advantage of DAP-seq is that the influence of epigenetic DNA modifications (e.g. methylation) on TF binding can be assessed by generating a PCR amplified DNA library (ampDAP-seq) where all DNA modifications are removed. Comparison of both *Arabidopsis* libraries revealed that of the 529 TFs assayed with DAP-seq, more than 70% is sensitive for DNA methylation (O'Malley et al., 2016). In that regard, different tissues contain different DNA modifications and comparison of these tissue-specific genomes can further elucidate the epicistrome of an organism. Overall, this technique can be used for every organism for which the genome is sequenced and has recently been applied in bacteria, maize and Eucalyptus (Galli et al., 2018; Brown et al., 2019; Cao et al., 2020; Zhang et al., 2020).

Chromatin immunoprecipitation

Although one can easily predict which DNA binding motif is specific for a TF of interest with previous mentioned *in vitro* methods, it does not give any information about the endogenous regulating function of the TF. Therefore, additional *in vivo* analyses should be performed to obtain relevant interactomics data. The most well known and most used *in vivo* TF-centered method is chromatin immunoprecipitation (ChIP) (Park, 2009; Collas, 2010).

ChIP allows for the isolation of a TF together with its target sequences. The different steps of ChIP are presented in figure 5. First the *in vivo* interactions are fixated by

cross-linking (usually performed by formaldehyde) where after the cells are lysed and nuclei are isolated. Chromatin is fragmented via sonication and afterwards the TF is purified with the covalent bound target sequences via immunoprecipitation. Once the complexes are isolated, fixation of the interactions is reversed so that the target sequences can be purified and analyzed. Analysis of the purified DNA can be done on a genomic DNA-microarray (chip), although using a chip is completely outdated with the development of next-generation sequencing (NGS).

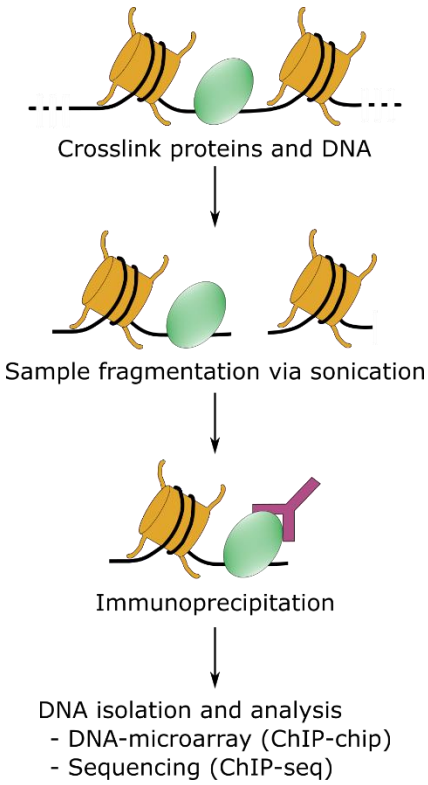


Figure 5: Schematic description of ChIP. *In vivo* interactions are fixated by cross-linking. Chromatin is fragmented via sonication and afterwards the TF is purified with the covalent bound target sequences via immunoprecipitation. Once the complexes are isolated, fixation of the interactions is reversed so that the target sequences can be purified and analyzed on a genomic DNA-microarray (chip) or via next-generation sequencing (NGS).

ChIP has one important aspect that allows to capture *in vivo* interactions; the fixation of protein-DNA interactions by cross-linking. This covalent binding allows for purification without losing the interaction during the process. Many cross-linking agents exist but only a few can be applied for ChIP. It is important that the cross-linking agent is able to fixate protein-DNA interactions, and that this fixation can be reversed to allow the isolation of the bound DNA sequence. Furthermore, the cross-linking should not have a negative effect on the efficient purification of the protein-DNA complex. Formaldehyde (CH_2O) is one of the few cross-linking agents that meets these requirements. It fixates protein-DNA interactions but also protein-protein and protein-RNA interactions, as long as these macromolecules are in close proximity of within 2Å. Formaldehyde reacts with amino and imino groups of amino acids (lysine, arginine and

histidine) and DNA. This results in an intermediate Schiff base that further reacts with a second amino group to form the final condensed covalent binding (figure 6). To reverse the cross-linking one can simply heat up the sample.

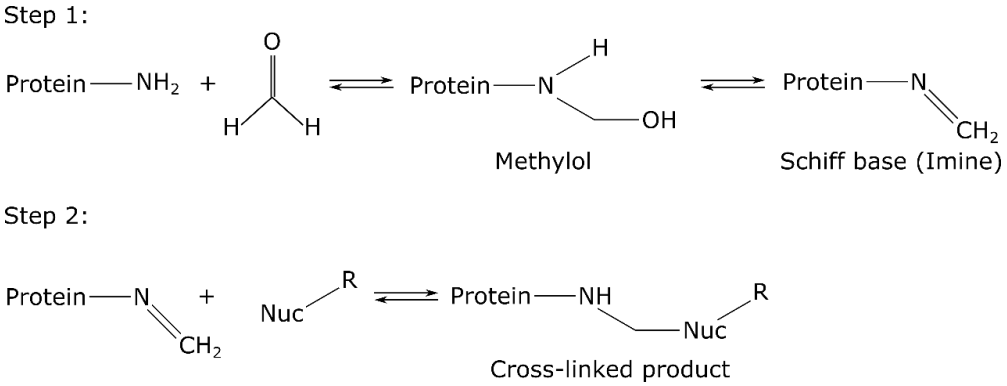


Figure 6: Formaldehyde cross-linking reaction. Formaldehyde reacts with amino and imino groups of amino acids. This results in an intermediate Schiff base that further reacts with a second amino group (e.g. of DNA) to form the final covalent cross-linked product.

To perform a successful ChIP experiment, a highly specific antibody against the protein of interest is needed. The antibody must be able to recognize the target protein in its native chromatin-associated context after fixation. Alternatively, affinity purification can be performed upon overexpression of a tagged TF. Another technical challenge to obtain a proper NGS analysis, is the requirement of a DNA yield around 10 ng, which highly depends on the DBP. For example, when analyzing a common DBP, like a histone (or histone modification), a lower amount of input material is needed compared to a sequence-specific TF, to obtain similar DNA yields. Contamination of unbound DNA sequences during immunoprecipitation is one more obstacle that leads to the identification of false positives and missing out lower affinity sites (false negatives). Therefore, an alternative ChIP-seq approach has been developed implementing strand-specific 5′-3′ degradation by lambda exonuclease (ChIP-exo) (Rhee and Pugh, 2011; Matteau and Rodrigue, 2015; Rossi et al., 2018). The exonuclease removes non-cross-linked DNA and trims immunoprecipitated DNA sequences up to the cross-linked TF. As a result, ChIP-exo provides a genome-wide and high resolution set of TF binding sites.

ChIP is the standard method for genome-wide detection of binding regions of TFs. In principal ChIP-seq can be applied to all species of which the genome sequence is completed. Plant ChIP experiments are mainly performed in *Arabidopsis* and different protocols have been published (Kaufmann et al., 2010; Zhu et al., 2012). Deng et al. identified 505 target genes of the flowering locus C (FLC) with ChIP-seq and determined the consensus sequence to be CCAAAAAT(A/G)G in *Arabidopsis* (Deng et al., 2011). In *Zea mays* ChIP-seq data combined with RNA-seq data revealed 35

opaque 2 (O2)-modulated target genes containing 4 different binding motifs (Li et al., 2015), and also for rice a ChIP-seq protocol exists (Zhu et al., 2012). ChIP experiments in tree species still remains a major challenge because of their thick cell wall layer. But in 2013 Lin *et al.* established a robust ChIP assay using *Populus trichocarpa* protoplasts from stem-differentiating xylem to validate a TF-DNA interaction involved in wood formation (Lin et al., 2013; Li et al., 2014).

Tandem chromatin affinity purification

As mentioned before, a successful ChIP-seq experiment relies on the efficient enrichment of DNA target regions using highly specific antibodies. However, ChIP-grade antibodies are not available for every DBP. Therefore, one-step affinity purification has been applied using a tagged version of the DBP of interest. Although successful one-step purifications have been performed (Kaufmann et al., 2010; Deng et al., 2011; Li et al., 2015), the signal-to-noise ratio is rather low. To increase signal-to-noise ratios, tags with very high affinity, such as *in vivo* biotinylation, were tested (Viens et al., 2004; van Werven and Timmers, 2006). The binding between biotin and streptavidin represents one of the strongest non-covalent interactions known ($Kd = 10^{-15}$ M). The superior strength of the biotin–avidin interaction allows for more stringent washing conditions during the ChIP protocol, resulting in a better signal-to-noise ratio.

However, biotin is an essential cofactor in plants and histones are naturally biotinylated, making it a suboptimal tag for TF localization studies in plants. Therefore, our research group developed a tandem chromatin affinity purification technique (TChAP) leading to reduced nonspecific biotin background signals and retaining the stringent washing step during the biotin-avidin interaction (Verkest et al., 2014). An added advantage of the stringent affinity purification is that contamination of plastid DNA is lacking which makes nuclei isolation unnecessary.

The tag contains a biotinylated peptide flanked by a repeat of 6 times histidine (HBH) and is compatible with formaldehyde cross-linking. As in a standard ChIP protocol the cross-linked cells are lysed and chromatin is sheared via sonication. A first affinity purification step is performed with a nickel-charged affinity resin to capture the polyhistidine sequences. After several washing steps, removing the endogenous biotinylated proteins, the affinity of micromolar range is then relinquished by adding an excess of imidazole. The second affinity step uses streptavidin-sepharose beads. More stringent washing steps can be applied here to obtain a very pure DNA yield for subsequent NGS.

As a proof of principle, the E2Fa TF activator was analyzed via TChAP. E2Fa is a key regulator in the G1-S phase transition during cell cycle in plants (Inze and De Veylder, 2006). The E2Fa DNA binding motif determined by Vandepoele in 2005 (TTTCCCGC) is overrepresented in cell cycle, DNA replication, DNA repair and chromatin structural genes like origin of replication complex 1b (*ORC1b*), chromosome transmission fidelity 18 (*CTF18*), and E2F target gene 1 (*ETG1*) (Vandepoele et al., 2005). By performing TChAP on the E2Fa TF we were able to identify 42% of the E2Fa-regulated genes. Although the yield was low compared to ChIP and single step chromatin affinity purification (ChAP), TChAP showed the highest enrichment of E2Fa regulated genes, indicating a high signal-to-noise ratio. To obtain enough DNA yield for sequencing, starting cell material should be 3-fold higher than for ChIP. Applying TChAP in cell suspension culture ensures for an unlimited supply of cells, but optimizations should be considered when switching to plants.

ChIP with selective isolation of chromatin-associated proteins

Another alternative for ChIP-seq is ChIP with selective isolation of chromatin-associated proteins (ChIP-SICAP) (Rafiee et al., 2016). Similar to TChAP, ChIP-SICAP uses a two-step purification of a protein of interest to obtain a high signal-to-noise ratio. Also here the strong interaction between biotin and avidin is used to obtain a highly purified protein-DNA complex. The difference however lies within the two purification steps. Both steps have a different target, meaning that during the first purification step the protein of interest is immuno-precipitated, while the second purification steps targets the bound DNA which is end labeled with biotin. Figure 7 displays an overview of the different ChIP-SICAP steps.

ChIP-SICAP is an efficient tool for effectively removing common contaminants allowing for high-quality NGS and at the same time MS to identify chromatin-bound partners of the bait protein. However, the need for a ChIP-grade antibody during the first affinity step creates the limitation of analyzing only well-known chromatin associated proteins. Also the two-step purification ensures for a lower yield whereby higher input material is necessary and very low-abundant proteins will not be detected with MS.

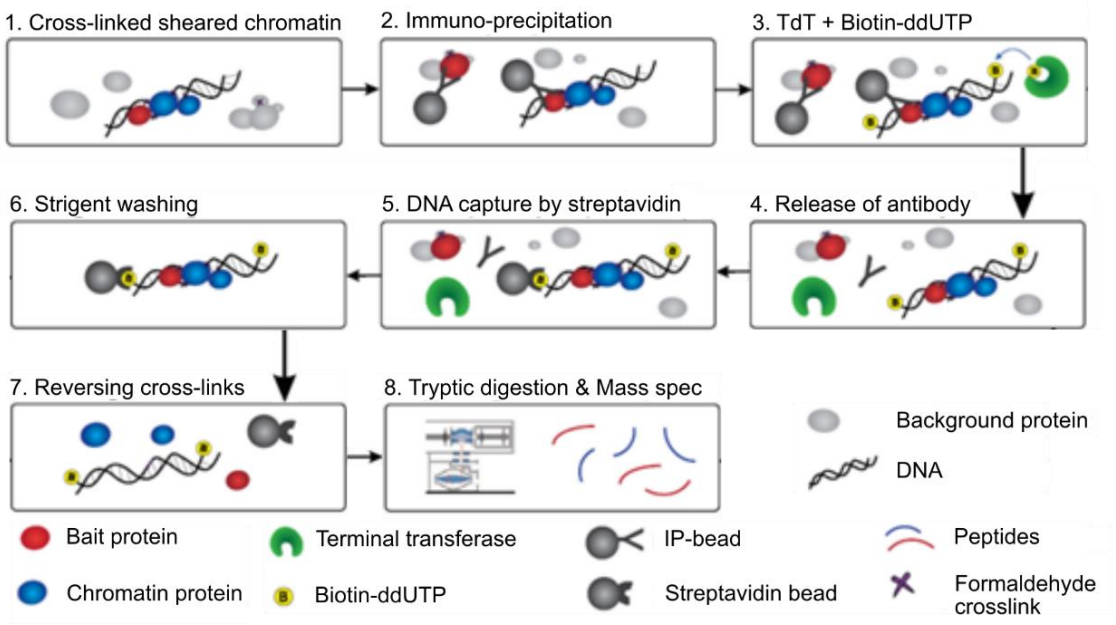


Figure 7: Schematic representation of ChIP-SICAP. Similar to a ChIP experiment, DNA proteins are crosslinked by formaldehyde, and fixed chromatin is sheared to small fragments by sonication (1). Following immunoprecipitation with a suitable antibody (2), DNA is biotinylated by TdT and biotin-ddUTP (3). The antibody is denatured by SDS (4), and chromatin is retrieved along with interacting proteins on streptavidin beads (5). Following extensive washing (6), isolated chromatin fragments are heated to reverse the crosslinks (7). Finally, proteins are digested and identified by mass spectrometry (8).Figure adapted from Rafiee et al., 2016.

Electrophoretic mobility shift assay

To identify the DNA binding motif and target genes of a specific TF, previously mentioned techniques can be applied. Once identified, these interactions should be validated by a second, independent TF-centered method or by an electrophoretic mobility shift assay (EMSA). This technique is based on a simple concept: 'a protein-DNA complex will migrate slower during electrophoresis then the DNA sequence itself' (Lane et al., 1992; Schwechheimer et al., 1998; Hellman and Fried, 2007). Once the protein-DNA interaction is identified, the recombinant protein can be mixed with a pool of labeled target DNA sequences, followed by electrophoresis. Migration retardation of the labeled DNA will occur when there is binding between the two and thus a shift to higher molecular weight. As a negative control, only the labeled DNA sequence is loaded on the gel. Migration retardation can further be influenced by the number of proteins that bind one DNA sequence, the protein charge and the bending of the DNA during protein binding.

EMSA can also be applied to determine the essential base pairs for binding upon a target sequence by analyzing mutated target sequences. When essential base pairs are mutated, binding will not occur, leading to a loss of migration retardation. Schramm et al. affirmed in this way the binding of DREB1B and DREB2A/B on the *HsfA3* promoter and confirmed the presence of the corresponding DRE1 and DRE2 binding

elements in *Arabidopsis thaliana* (Schramm et al., 2008). Alternatively, one can execute a footprinting analysis as explained above.

Gene-centered tools

TF-centered techniques are ideal when defining the transcription regulatory mode of action of a specific TF. Many successful examples can be found back in literature. But when an unknown protein is causing a certain transcriptional outcome, the research possibilities abate. Instead of putting the TF or protein at the center, the focus shifts towards the transcriptionally regulated gene. The gene, and more importantly the cis regulatory elements, will serve as a bait to capture the binding proteins. To do so, different gene-centered methods have been developed and will be discussed in detail.

DNA affinity chromatography

A first approach is by immobilizing synthetic DNA strands on a solid support like sepharose or magnetic beads, followed by incubation with a cell lysate or nuclear extract (Gadgil et al., 2001; Lambert et al., 2012). The immobilized DNA strands work as affinity probes and will specifically enrich for associated proteins. Including several washing steps will diminish the nonspecific binding proteins, and after elution the interacting proteins can be identified via mass spectrometry. Despite the different washing steps, a lot of false positive proteins are being co-purified. To circumvent this problem, quantitative proteomics have been applied (Mittler et al., 2009), or the usage of nonspecific or mutated DNA sequences (Gadgil et al., 2001; Kwon and Chung, 2004).

In *Arabidopsis thaliana* an interactor of the single-stranded telomeric repeat was identified using DNA affinity chromatography (Kwon and Chung, 2004). A single stranded $(TTTAGGG)_6$ oligodeoxyribonucleotide was used as a bait and incubated with an *Arabidopsis* extract. To lower the background proteins, the extract was pre-incubated with nonspecific single-stranded oligonucleotides. After elution in SDS-PAGE sample buffer, a MALDI-TOF MS analysis was performed identifying the single-stranded telomere binding protein 1 (STEP1). They also showed via EMSA that STEP1 is a plant specific binder of the single-stranded telomeric repeat, lacking any binding with human and *C. elegans* single-stranded telomeric repeats and the double stranded plant telomere sequence. In 2012 DNA affinity chromatography has been applied to identify TFs responsible for the activation of the high light and redox sensitive LHCB2.4 gene (At3g22840) (Shaikhali et al., 2012). A 144-bp promoter fragment was used to capture interacting TFs from high light-treated plants. Subsequent Q-TOF MS analysis

revealed the binding of the bZIP16 TF on the promoter fragment, which was validated by EMSA.

Protein microarray

Next to the protein binding microarrays where a DNA microarray is used to identify target sequences of a specific protein, protein microarrays exist to reveal interaction with a specific DNA sequence. A microarray containing a hundred or thousands of proteins can be applied to identify DBP for a specific labeled DNA sequence (Chen and Zhu, 2006; Tao et al., 2007; Xie et al., 2011). Similar to PBM, unbound DNA is removed by washing steps and bound DNA is visualized with a fluorophore like CY3 (figure 8). A Cyc5-labeled mutant DNA probe can be used to correct for non-specific biding.

Protein microarrays can also be used to analyze protein-RNA, protein-lipid, protein-metabolite, protein-antibody and protein-protein interactions. Identifying substrates of enzymes via protein microarrays as well belongs to the possibilities (Zhu and Snyder, 2003). The first *Arabidopsis* protein microarray was developed by Gong et al. and contained 802 TF's (Gong et al., 2008). He demonstrated the utility for identification of TF-DNA and TF-protein interactions. 49 new TFs were identified belonging to the AP2/ERF family binding the GCC box and DRE element. They also identified four previously unknown binding proteins of the HY5 TF which were validated by a yeast two-hybrid (Y2H) assay.

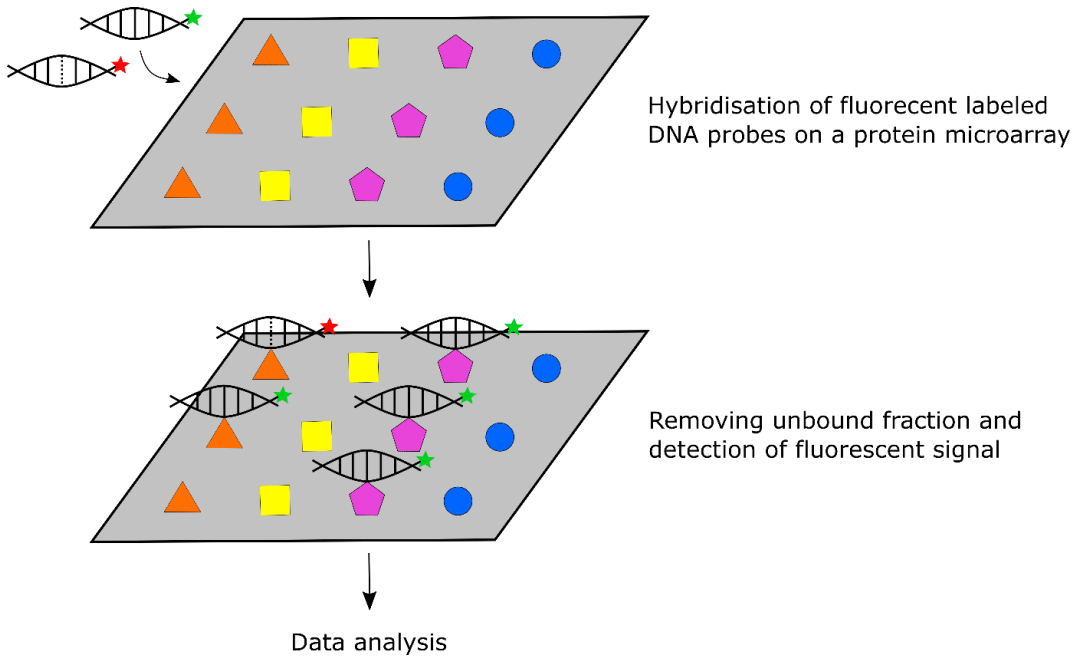


Figure 8: Schematic description of protein microarray. Green star = Cy3, Red star = Cy5. Figure adapted from Xie et al., 2011.

Yeast one-hybrid

The most well-known and most applied gene centered technique for plant protein-DNA interactions is yeast one-hybrid (Y1H) (Deplancke et al., 2004; Reece-Hoyes and Marian Walhout, 2012). This technique includes two components. The first component is the bait construct containing the DNA sequence of interest, mostly repeats of a cis regulatory element (CRE) or a short promoter sequence instead of a full promoter sequence. This DNA bait is followed by a gene encoding for a reporter protein. The second component is an expression construct encoding for a prey TF fused to a yeast transcription activator domain (AD). Binding of the prey TF on the bait DNA will lead to transcription of the reporter and hence the detection of a positive interaction (figure 9).

The most frequently used reporter genes are *HIS3* and beta-galactosidase. HIS3 is a component of the histidine biosynthesis pathway, enabling yeast to grow in the absence of exogenously supplied histidine. When *HIS3* is expressed in *his3* mutant yeast, colonies will grow on medium lacking histidine, indicating an interaction between the TF and DNA bait. This is in contrast with the beta-galactosidase reporter where all colonies will grow regardless of the interaction, but need to be tested colorimetrically. Upon interaction, beta-galactosidase will convert the colorless X-gal to 5,5'-dibromo-4,4'-dichloro-indigo, a blue compound. This makes *HIS3* an easier option as a positive reporter for high-throughput screens. Usually both positive reporters are combined in a Y1H screen, allowing a first selection for *HIS3* expression and followed by a colorimetric test with X-gal. Only colonies that score positive for both reporters are considered for further analysis. Also auto-activation can occur, where the bait report construct is activated on its own in yeast. Yeast strains without prey construct can be assayed for auto-activation using the positive reporter HIS3 and its inhibitor 3AT. In this case, auto-activation is attenuated by adding 3AT. Subsequently, protein-DNA interactions that activate reporter expression at a level that is higher than the auto-activity will eventually lead to growth on medium lacking histidine. Alternatively, one can use URA3 as negative and positive reporter system (Yanai, 2013). In a first step, yeast strains only containing the bait construct are plated on a selective medium containing 5-fluoro-orotic acid (5-FOA). When there is auto-activation, the protein product of URA3 (Orotidine 5'-phosphate decarboxylase) will convert 5-FOA into a toxic compound, killing all yeast cells with auto-activating bait construct. Consequently, only non-auto-activating yeast strains remain for transformation with prey constructs. Upon binding, URA3 expression will allow growth on medium lacking uracil. (figure 9).

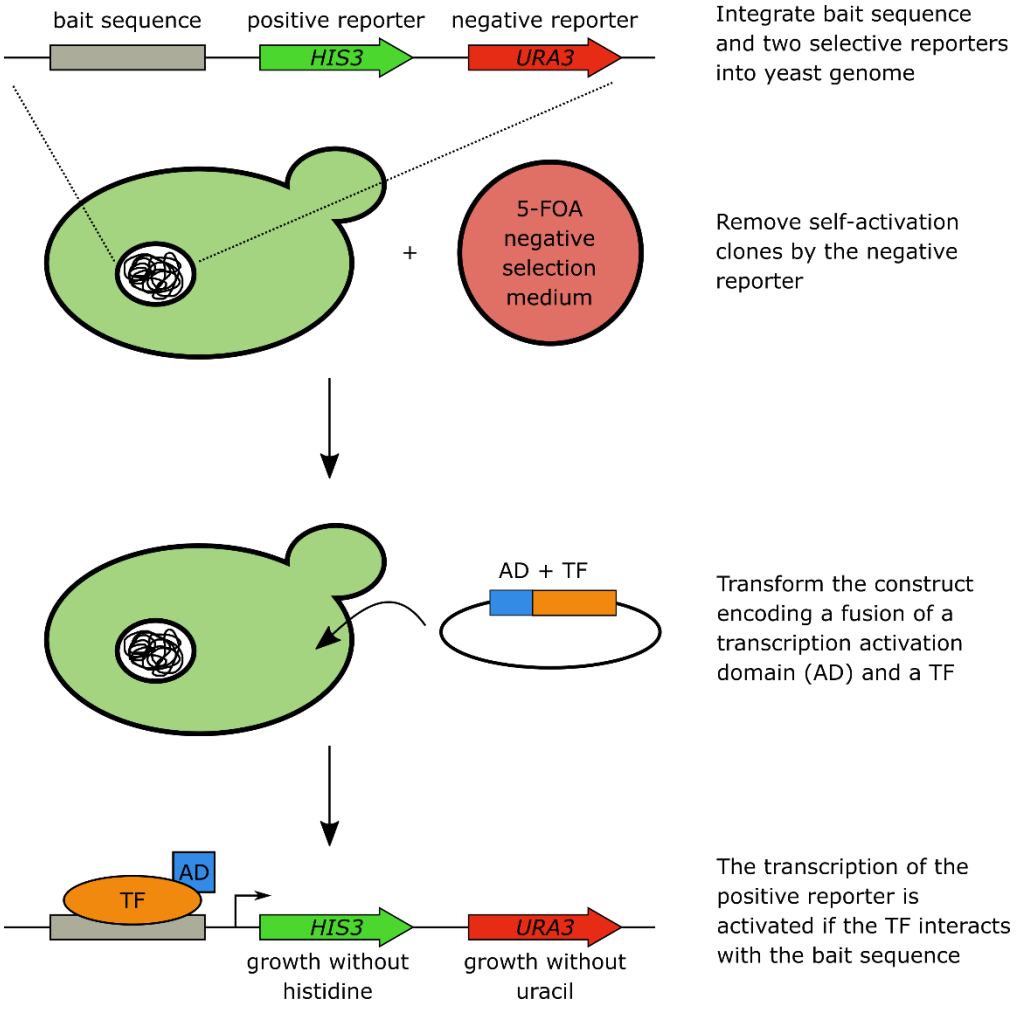


Figure 9: Workflow of yeast one-hybrid (Y1H). Figure adapted from Xie et al., 2011.

For the prey constructs one can choose to screen a cDNA library or a TF library. Although cDNA libraries are available for almost all organisms, these are poor candidates to assess protein-DNA interactions in a Y1H screen. TFs account for only 5% to 10% of an organism's protein-coding genes and they are generally expressed at low level. Therefore, cDNA libraries encode a very low amount of TFs, making the Y1H screen inefficient and time consuming. TF libraries would be a better option but making them is not straightforward. Based on the presence of structural protein domains and amino acid sequence, potential TFs are identified from the proteome and the corresponding open reading frame has to be individually cloned from a cDNA source. Though TF libraries are available for some organisms, including *Arabidopsis thaliana* and maize (Mitsuda et al., 2010; Castrillo et al., 2011; Burdo et al., 2014), these libraries are far from complete and may include incorrect ORFs due to a suboptimal TF prediction strategy. In recent years robot-assisted screens have been developed where arrays (e.g. 96 well plates) of prey constructs are being tested (Gaudinier et al., 2011; Reece-Hoyes et al., 2011; Taylor-Teeples et al., 2015). The

advantage is that it is less time consuming and there is no need for sequencing of positive clones as the identity of the prey construct per well is already known.

Prey proteins are fused to an AD, usually the yeast GAL4 AD including a nuclear localization signal (NLS). It was shown that the AD of GAL4 can be separated from its DNA binding domain and retain its function (Brent and Ptashne, 1985). When the GAL4 AD was fused to other proteins, the activating function remained even when it was fused to a repressor (Ma and Ptashne, 1988). This however means that a Y1H screen can only determine the binding capacity of a TF for a specific DNA sequence but the regulatory function of the TF cannot be identified. Another disadvantage is that this is a binary system, identifying interaction of one protein at a time. This means that TFs that work as multimers are not detected in a Y1H screen, nor the whole regulatory complex of a specific promoter. Furthermore, Y1H screens for plant protein-DNA interactions are outside of the endogenous context. Protein-DNA interactions which are dependent on plant specific post translational modification will therefore not be detected.

Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics

To be able to identify *in vivo* protein-DNA interactions in a gene centered way, different groups have developed new approaches where specific DNA is captured via hybridization. One of these methods is Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics (GENECAPP), illustrated in figure 10 (Smith et al., 2011; Wu et al., 2011). Similar to ChIP, formaldehyde is used to fixate protein-DNA interactions *in vivo*, after which the chromatin is fragmented. Subsequently, an exonuclease is added to remove one of the two DNA strands of the protruding ends. The free single-stranded DNA region is then hybridized with complementary single-stranded DNA of interest on a solid support (e.g. a chip or beads). Non hybridizing protein-DNA complexes are washed away after which standard mass spectrometry (MS) is applied to identify all the proteins that are bound to the DNA sequence of interest.

As a proof-of-principle Wu et al. (2011) identified the binding of FoxO1 on the *IGFBP1* promoter. However, this complex was formed *in vitro* and until now no *in vivo* interactions have been identified by GENECAP. They postulate that the biggest obstacle is the detection sensitivity of the MS causing the need for an enormous amount of starting material. TFs are present at a very low abundance, and in an ideal situation, two TFs per diploid cell would be isolated during this gene-centered protocol.

To obtain enough protein yield that is detectable via MS (which is at least a picomole of protein), an amount of 6 x 10^{11} cells is needed which resembles 100 liters of human cell culture medium. This implies that GENECAP is more suitable to study abundant DBPs likes histones or repetitive DNA sequences like telomeres and ribosomal DNA. But because of the growing advances in instrument sensitivity for MS, GENECAP might evolve as a valuable *in vivo* PDI discovery technique in the coming years.

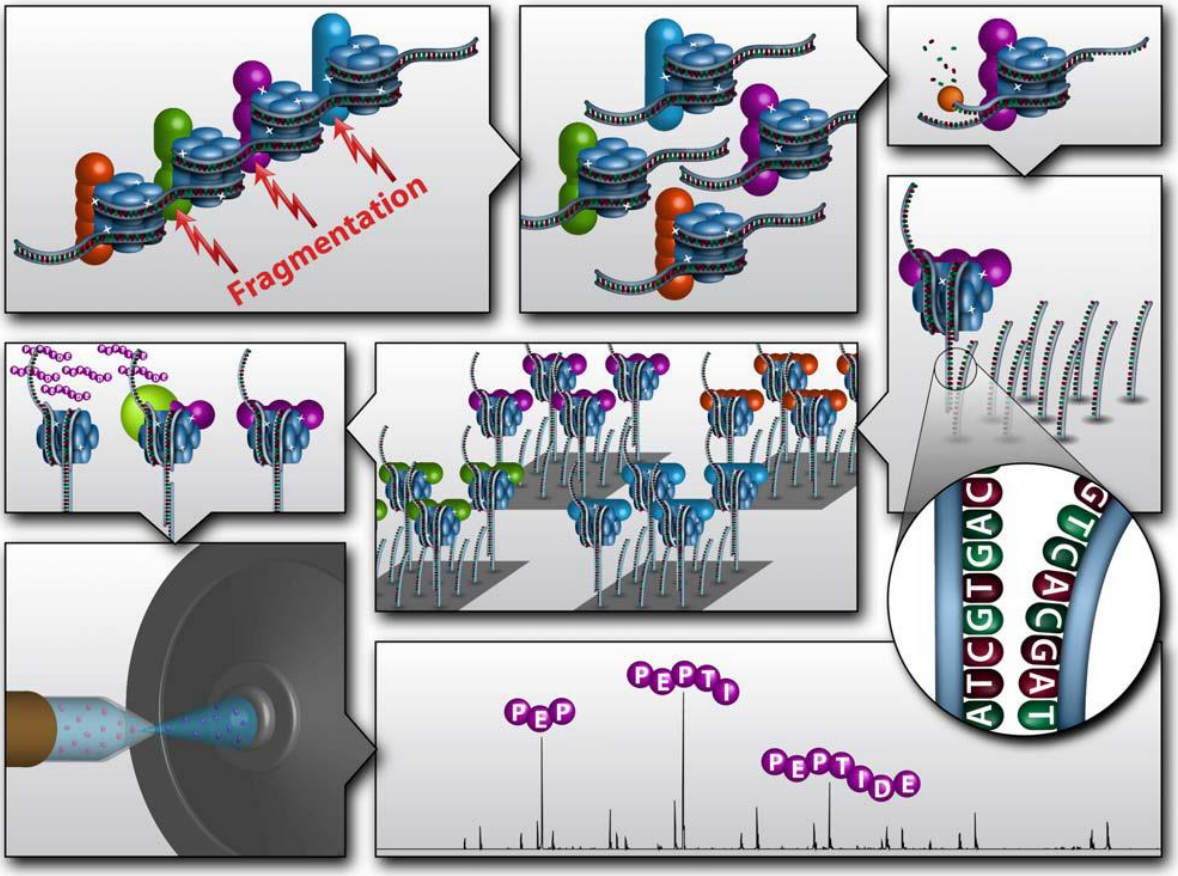


Figure 10: Schematic diagram of GENECAPP, Global ExoNuclease-based Enrichment of Chromatin-Associated Proteins for Proteomics. In this illustration of the process, formaldehyde cross-linked chromatin is fragmented (e.g. sonication or restriction endonuclease) into small, nucleosome-length pieces. Fragments are treated with exonuclease to produce single-stranded regions, which are used for sequence-specific capture on a complementary DNA oligonucleotide array. Protease digestion of the captured complexes yields sample for MS analysis; enabling identification of the proteins and subsequent association of those proteins with genomic loci. Figure adapted from Smith et al., 2011.

Proteomics of isolated chromatin segments

Another gene-centered approach that uses a DNA probe to isolate specific sequences *in vivo* is proteomics of isolated chromatin segments (PICh) (Dejardin and Kingston, 2009). Dejardin et al. (2009) took into account the MS sensitivity limitation and focused on the identification of telomere specific proteins as a proof-of-principle. Also here formaldehyde is used to fixate the *in vivo* interactions and sonication is applied to fragment the chromatin. Next, a specific DNA probe against the chromatin region of interest is added, linked to an affinity tag via a spacer arm. This probe contains locked nucleic acid (LNA) residues, which have a different structure as the common nucleic acids, leading to a stronger binding affinity with the complementary DNA string (Vester and Wengel, 2004). As affinity tag an analogue of biotin, desthiobiotin, is used so that the specific protein DNA complex can be isolated with magnetic streptavidin beads. By adding biotin, which has a higher affinity for streptavidin compared to desthiobiotin, the isolated complexes are eluted. Purified interactors are subsequently identified via MS. Figure 11 displays an overview of the different PICh steps.

By using a DNA probe against human telomere sequences, Dejardin could identify 33 well known telomere binding proteins like APOLLO, TRF1 and TRF2, but also unknown telomere interactors were detected and validated via localization studies. As a control he used a DNA probe with the same base composition but in a scrambled order (Dejardin and Kingston, 2009). PICh has also been applied to define the protein landscape of the *Drosophila* Telomere-Associated Sequence (TAS) repeats which are subtelomeric regions of chromosomes 2, 3, and X. Here 70 potential interactors were identified of which 5 were validated by ChIP (Antao et al., 2012). Successful results were also obtained a few years later when comparing functional human telomeres with dysfunctional telomeres (Bartocci et al., 2014). Changes in chromatin composition could be visualized by applying PICh, showing the recruitment of DNA damage response (DDR) factors to the dysfunctional telomeres. Not only telomere regions have been analyzed by PICh, also interactors of the promoter region of the ribosomal RNA genes (Ide and Dejardin, 2015) and barley centromeric chromatin (Zeng and Jiang, 2016) have been identified via PICh.

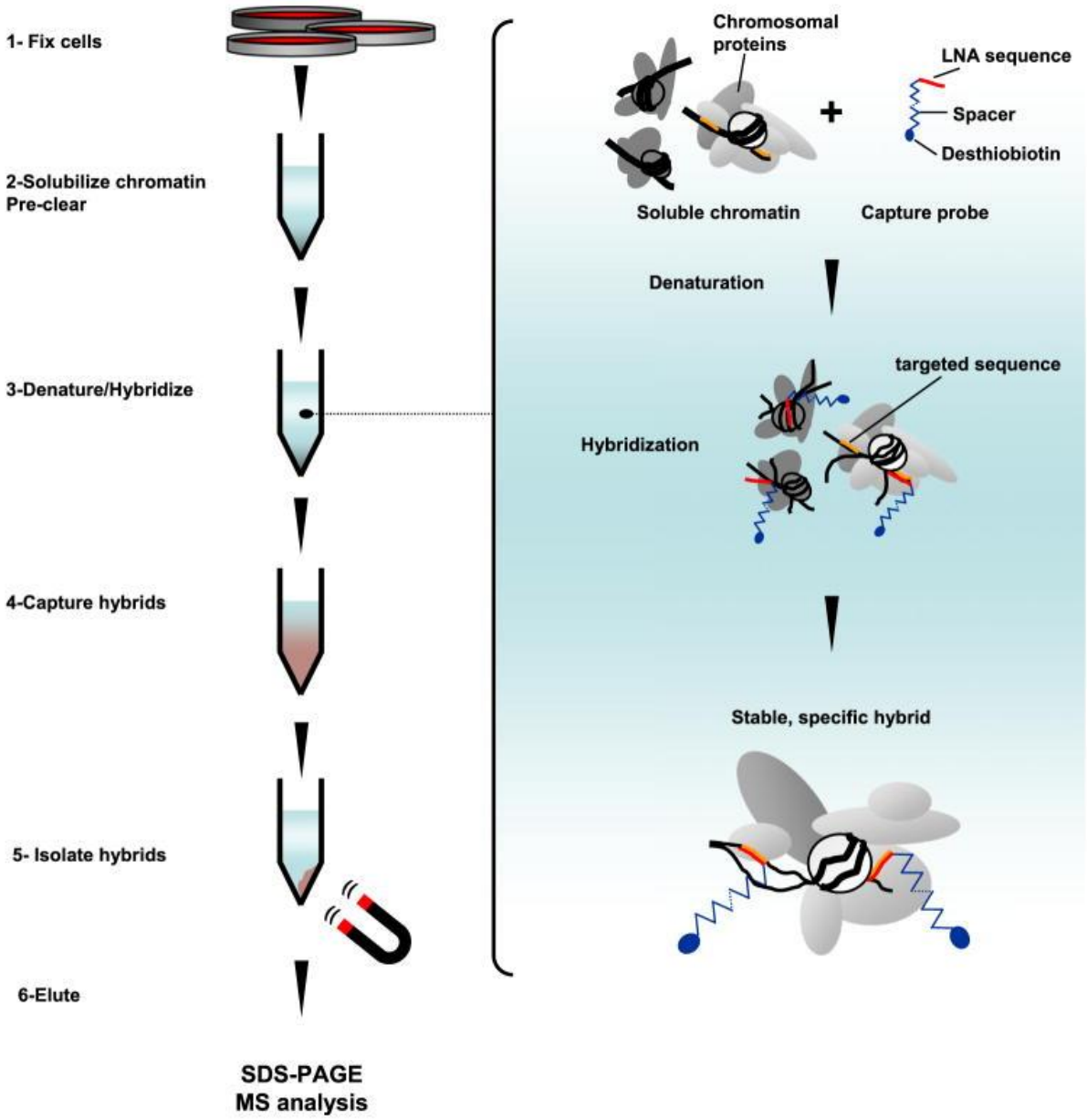


Figure 11: Workflow of proteomics of isolated chromatin segments (PICh). LNA sequence: DNA probe with locked nucleic acid residues. Figure adopted from Dejardin and Kingston, 2009.

Insertional chromatin immunoprecipitation

Although PICh has been applied successfully in different species (human cells, Drosophila and barley), the limitation of MS sensitivity remains. PICh isolation of single copy loci has not been performed yet because of the large amount of starting material that is needed to obtain enough protein yield for MS. Other groups however have developed gene-centered tools to isolate specific genomic regions. One of these techniques is called insertional chromatin immunoprecipitation (iChIP) (Hoshino and Fujii, 2009). While GENECAP and PICh make use of DNA hybridization, a genetic engineering step is implemented in iChIP for isolation of specific DNA sequences. A

plasmid containing the genomic region of interest in proximity of a repeat of an exogenous DNA binding sequence is transfected in a cell line expressing the complementary exogenous DBP fused to an affinity tag. In the case of iChIP, an 8 times repeat of the LexA binding element (LexA BE) is placed nearby a promoter sequence and the LexA DB domain is expressed fused to a nucleus localization sequence (NLS) and FLAG affinity tag. To allow for tandem purification, the affinity tag includes a tobacco etch virus (TEV) protease cleavage site and calmodulin-binding peptide sequence. Because LexA binds in the region of interest, one can now easily do affinity purification in analogy with ChIP. Also here, cross-linking is applied for fixation and chromatin is fragmented to a certain bp length. After purification of the desired complex, reverse cross-linking is done to be able to identify the isolated proteins (figure 12). Also interacting genomic regions like distal enhancers or intra- and interchromosomal interactions can be detected via DNA microarray's or NGS. Interacting RNA molecules can be identified using a reverse transcription polymerase chain reaction (RT-PCR).

As a proof-of-principle Hoshino and Fujii tested the isolation of the *IRF-1* promoter and the identification of the interacting TF STAT1 (Hoshino and Fujii, 2009). Binding of STAT1 and expression of *IRF-1* was shown to be induced by interferon (IFN) γ, a signaling protein during immune responses in mammalian cells (Sims et al., 1993; Meraz et al., 1996). After immunoprecipitation with anti-FLAG, Hoshino and Fujii could show an enrichment of the specific genomic region with qPCR and the specific isolation of STAT1 upon IFNγ treatment. Without IFNγ induction the *IRF-1* promoter could still be isolated without any detection of STAT1.

A second iChIP analysis was performed on the chicken insulator HS4 (cHS4), which regulates expression of β-globin genes in mammalian cells (Fujita and Fujii, 2011). Insulators protect nearby genes from undesirable enhancers as well as from chromatin silencing, but their regulation and precise molecular mechanism is somewhat unclear. By using a construct containing two 8 times LexA binding sequence repeats both flanked by a 6 times repeat of cHS4, different proteins and RNA molecules could be identified. By optimizing the affinity tag they could increase the affinity purification efficiency for cHS4 by 4-fold (Fujita and Fujii, 2012). This new tag contains 3 times FLAG, an NLS, DB and the dimerization domain of the LexA protein. They removed the TEV protease cleavage site and calmodulin-binding peptide because of the low TEV cleavage efficiency in cross-linked chromatin. Higher immunoprecipitation efficiency made the isolation of endogenous single copy loci possible. In 2015 iChIP was performed on the Pax5 1A endogenous promoter by inserting the 8 times LexA BE in proximity via homologous recombination in a chicken B cell line (Fujita et al.,

2015). In combination with quantitative (stable isotope labeling by amino acids in cell culture, SILAC) MS analysis they identified the Thy28 TF as a regulator of Pax5 1A promoter.

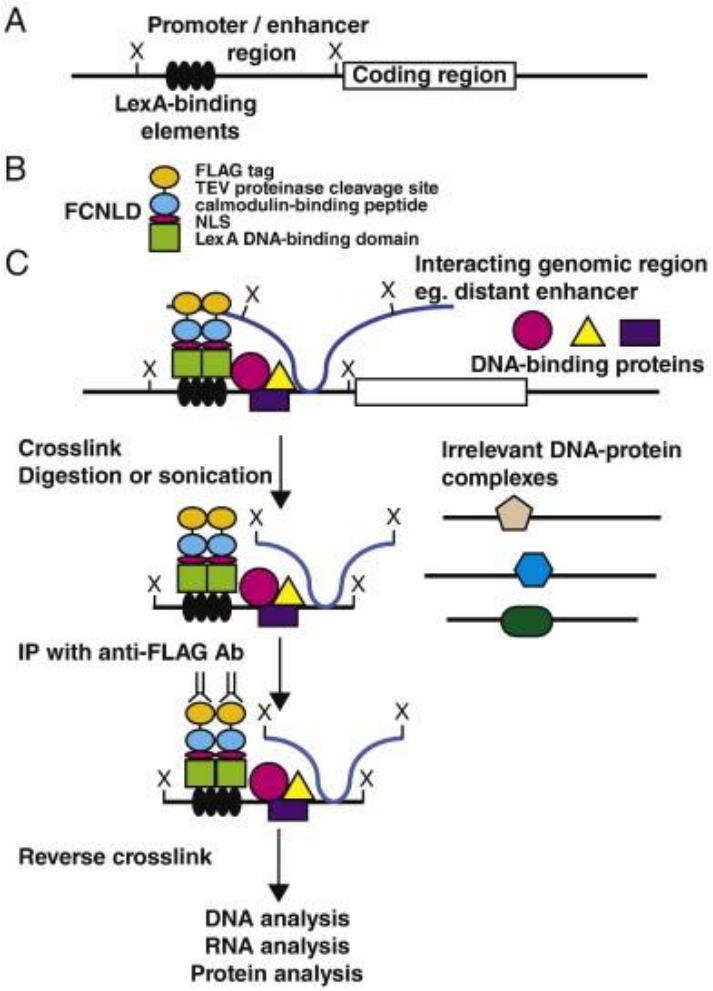


Figure 12: Scheme of insertional chromatin immunoprecipitation (iChIP). The system consists of a promoter/enhancer element of a gene of interest linked to LexA-binding sites (the LexA-tagged promoter) **(A)**, and FLAG-tagged, nuclear localization signal (NLS)-fused LexA DNA-binding domain (FCNLD) **(B)**. A TEV protease cleavage site and calmodulin-binding peptide sequence are fused to allow tandem purification scheme. Cells expressing FCNLD are transiently or stably transfected with the LexA-tagged promoter. Alternatively, LexA-binding sites are knocked-in in the promoter/enhancer element of the gene of interest in cells expressing FCNLD. These cells are stimulated with ligand of interest, cross-linked with formaldehyde, and lysed. Then, cross-linked DNA is digested with a restriction enzyme or fragmented by sonication. Subsequently, the LexA-tagged promoter is immunoprecipitated with anti-FLAG antibody, and crosslink is reversed. Molecules (DNA, RNA, proteins, and others) associated with the LexA-tagged promoter are isolated and characterized **(C)**. Figure adapted from Hoshino and Fujii, 2009.

Chromatin affinity purification with mass spectrometry

In yeast, a similar approach like iChIP was developed called chromatin affinity purification with mass spectrometry (ChAP-MS) (Byrum et al., 2012). Also here a genetic engineering step is implemented to insert a LexA DE in the DNA region of interest via homologous recombination, and protein A (ProtA) is fused to LexA to allow affinity purification with IgG resins. Figure 13 displays an overview of the ChAP-MS technique.

Byrum et al. placed the LexA DE upstream of the *GAL1* start codon to investigate *GAL1* regulation upon galactose activation and glucose inhibition. To enrich for bona fide interactors, a quantitative analysis was performed which involves a yeast strain lacking the LexA DE cultured with heavy isotopes ($^{13}C_6$ $^{15}N_2$-lysine), while the LexA DE

containing strain was cultured with light isotopes ($^{12}C_6$ $^{14}N_2$-lysine). By mixing both strains in a 1-to-1 ratio, non-specific proteins could be detected as they show a 1-to-1 ratio of heavy and light isotope-containing peptides during MS analysis. In this way, they could show that the inhibitory methylation of histone 3 on K36 (H3K36me3) was predominantly present during glucose treatment while acetylation motifs on H3 and H4 were more abundant during galactose activation. Also RNA polymerase and FACT (facilitates chromatin transcription) subunits were identified under active transcription of *GAL1*.

In 2017, ChAP-MS was applied in yeast to elucidate the double stranded break (DSB) repair mechanism at the *MAT* locus (Wang et al., 2017). several histone posttranslational modifications (PTMs) known to be involved in DSB repair were identified. More interestingly, 108 proteins were specifically enriched around the DSB, of which some have a general role in stress responses.

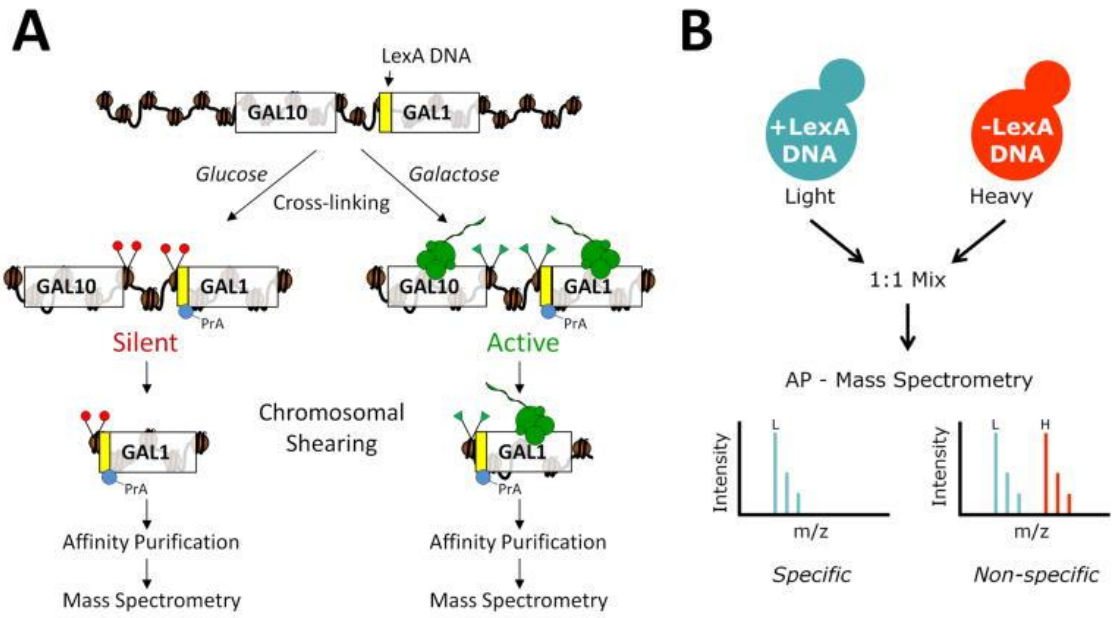


Figure 13: Overview of chromatin affinity purification with mass spectrometry (ChAP-MS). **A)** The chromatin affinity purification with mass spectrometry (ChAP-MS) approach provides for the specific enrichment of a given chromosome section and identification of specifically associated proteins and posttranslational modifications. A LexA DNA affinity tag was engineered just upstream of the GAL1 start codon in S. cerevisiae. Strains containing the LexA DNA binding site and a plasmid expressing LexA-PrA protein affinity tag were cultured in glucose or galactose to provide transcriptional repression or activation, respectively, and subjected to in vivo chemical cross-linking to trap protein interactions. Following shearing of the chromatin to 1,000 bp, LexA-PrA was affinity purified on IgG-coated Dynabeads and coenriched proteins/posttranslational modifications were identified by high-resolution mass spectrometry. **B)** To control for nonspecifically enriched proteins, a strain lacking the LexA DNA binding site, but containing the LexA-PrA plasmid, was cultured isotopically heavy ($^{13}C_6$ $^{15}N_2$-lysine) in glucose or galactose and mixed equally with the corresponding isotopically light culture containing the LexA DNA binding site prior to cell lysis. Following affinity purification (AP) and mass spectrometric analysis, nonspecifically enriched proteins were identified as a 1:1 ratio of light to heavy lysine-containing peptides, while proteins specifically enriched with the chromosome section were identified with a higher level of isotopically light lysine containing peptides. Figure adapted from Byrum et al., 2012.

TAL and CRISPR-based affinity purification

With iChIP and ChAP-MS it is feasible to identify interacting proteins and RNA *in vivo* for a specific single locus. However, some reservation must be made. Applying genetic engineering by inserting transgenes and making artificial repeats of the sequence of interest will change the chromosomal condition compared to the endogenous wild type locus. The obtained results are rather predictions and should always be validated by other techniques. Targeting the endogenous locus via homologous recombination of the LexA DE, as has been done for the Pax5 1A promoter and for ChAP-MS, results in a more reliable and a more complete list of interactors. But also here, the insertion of LexA DE could change the physiological chromatin structure and affect the binding of proteins. Therefore, new approaches have been developed that are able to isolate a specific DNA locus without inserting a transgene sequence repeat like LexA DE.

The discovery of engineered DNA-binding molecules such as zinc finger proteins, transcription activator-like (TAL) proteins and the clustered regularly interspaced short palindromic repeats (CRISPR) system, enabled new approaches for specific genome editing, transcriptional regulation, genomic loci visualization and isolation of genomic regions (Harrison et al., 2014). In 2013, the researchers who developed iChIP exchanged the LexA DE/LexA-NLS-3xFLAG for a TAL-NLS-3xFLAG that specifically recognize the telomere repeats (Fujita et al., 2013) (figure 14b). This new tool is called engineered DNA binding molecule-mediated chromatin immunoprecipitation (enChIP). Compared to iChIP, no insertion of exogenous DNA binding elements is needed, only the transgenic expression of a specific TAL protein fused to an affinity tag. In this way the endogenous genomic region of interest can be targeted without disturbing the chromatin structure and protein binding.

Specific isolation of telomere sequences with enChIP was successful and led to the identification of known and novel interacting proteins (Fujita et al., 2013). By combining enChIP with RNA-sequencing (RNA-seq), non-coding RNAs that interact with telomeres could be identified, like the telomeric RNA TERRA (Fujita et al., 2015). A TAL variant for ChAP-MS in yeast was also published in 2013, analyzing the *GAL1* locus under active transcription (Byrum et al., 2013). Using a ProtA fused TAL protein, histone acetylation patterns and RNA polymerase components were identified, thereby confirming the previous ChAP-MS results.

With the discovery of CRISPR and the development of an inactive Cas9 (dCas9, D10A/H840A) protein, enChIP was again modified to implement the CRISPR system (Fujita and Fujii, 2013). Here a gRNA specific for the genomic region of interest is transiently transfected in a cell line together with a dCas9 version fused to the 3xFLAG-

tag (figure 14a). With this approach the single copy locus *IRF-1* in mammalian cells could be specifically isolated together with chromatin related proteins, like histones and RNA helicases. However, transcription regulating factors could not be identified. When switching to a stable expression of the CRISPR components via retroviral transduction in combination with SILAC (enChIP-SILAC), specific *IRF-1* associated proteins could be identified and validated (Fujita and Fujii, 2014). Also a CRISPR variant for the ChAP-MS protocol in yeast exists (Waldrip et al., 2014). Here the dCas9 protein is fused to the ProtA affinity tag for specific isolation of an endogenous genomic region using a gRNA. To obtain proof-of-principle, the *GAL1* locus was again targeted under transcriptionally active conditions, and label free quantitative MS analysis was performed to identify specifically bound proteins and histone PTMs. Similar to the ChAP-MS and TAL-ChAP-MS results, histone acetylations and FACT components were identified, confirming the applicability of CRISPR-ChAP-MS. CRISPR-based affinity purification has been applied several times to elucidate the chromatin landscape of different loci in yeast, bacteria and mammalian cells (Campbell et al., 2018; Fujita et al., 2018; Fujita et al., 2018, 2018; Hamidian et al., 2018; Han et al., 2018; West et al., 2019), and an overview is given in figure 18.

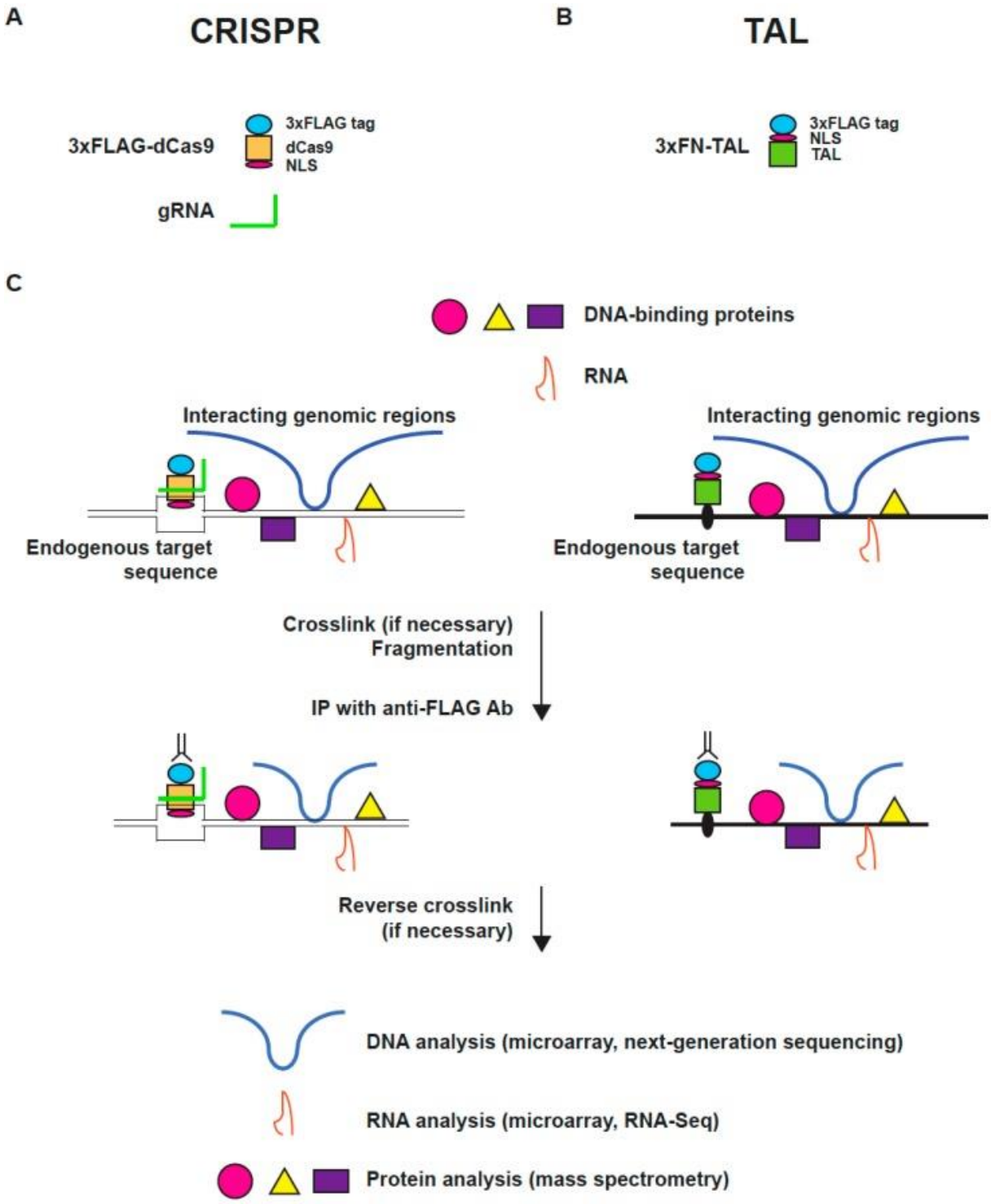


Figure 14: Scheme of enChIP. **A)** The locus-tagging CRISPR complex consists of tagged dCas9 (in this case, 3xFLAG-dCas9 consisting of the 3xFLAG tag, dCas9, and the nuclear localization signal (NLS) of SV40 T-antigen, as well as gRNA); **B)** The locus-tagging TAL protein (3xFN-TAL) consisting of the 3xFLAG tag, an NLS, and a TAL protein recognizing the target sequence; **C)** 3xFLAG-dCas9 and gRNA or 3xFN-TAL targeting the locus of interest are expressed in cells to be analyzed. The cells are cross-linked with formaldehyde or another crosslinker, if necessary, and chromatin is fragmented by sonication or enzymatic digestion. The tagged locus is isolated by affinity purification using anti-FLAG Ab. After reversal of cross-linking (when a crosslinker is used), molecules (proteins, RNAs, or other genomic regions) interacting with the target genomic region are identified by mass spectrometry, NGS, or other methods. Figure adopted from (Fujii and Fujita, 2015).

CRISPR affinity purification *in situ* of regulatory elements

In 2017 a CRISPR-based gene-centered approach was developed implementing the high affinity between biotin and streptavidin (Liu et al., 2017; Liu et al., 2018). This technique, called CAPTURE, makes use of a gRNA to direct a dCas9, fused to a FLAG and biotin-acceptor-site, to a specific genomic locus in mammalian cells (figure 15). These cells express the biotin-ligase BirA, which ensures *in vivo* biotinylation of dCas9. Subsequent streptavidin-based affinity purification of cross-linked cells allowed the researchers to capture telomeres together with known and unknown interacting proteins (Liu et al., 2017). With the same technique, a subset of single loci was isolated all involved in controlling five β-like globin genes in mammalian cells. Additionally, co-expression of different gRNAs resulted in simultaneous isolation of targeted single loci, demonstrating that CAPTURE can be adapted for multiplexed analysis of multiple loci. Combining CAPTURE with NGS does not only reveal enrichment efficiency of the target locus, but also evaluates the off-targets effects that may occur using the CRISPR system. Subsequent proteomics analysis led to the identification of known and new regulators of the five β-like globin genes (Liu et al., 2017). To correct for co-purification of endogenously biotinylated proteins and dCas9-associated non-specific proteins, a background list was composed based on purification in different control cell lines, including cells expressing only dCas9 or BirA. As mentioned before in the introduction, gene transcription is regulated by TFs that could be positioned at distal cis-regulatory elements like enhancers. Upon activation of transcription, these distal regions are recruited and will situate close to the promoter region. Liu et al. has adapted the CAPTURE technique in such a way to allow the identification of these distal DNA regions. Therefore, cross-linking was applied *in vivo* to fixate the distal chromatin regions, followed by CAPTURE and pair-end sequencing revealing several *de novo* CREs with unknown roles in globin gene regulation (Liu et al., 2017).

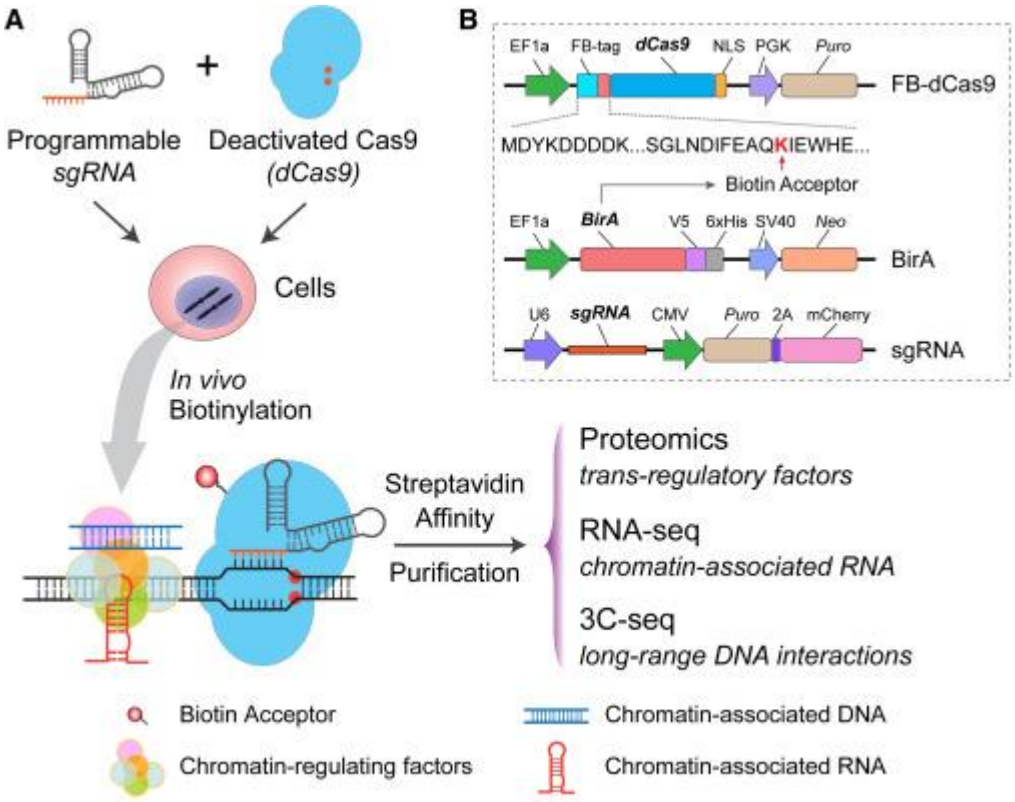


Figure 15: In Situ Capture of Locus-Specific Chromatin Interactions by Biotinylated dCas9. **A)** Scheme of dCas9-mediated capture of chromatin interactions. **B)** The three components of the CAPTURE system: a FB-dCas9, a biotin ligase BirA, and target-specific sgRNAs. Figure adopted from (Liu et al., 2017)

CRISPR-dependent biotin-based proximity labeling

Recent years, proximity labeling has become more and more popular for the investigation of interaction landscapes (Roux, 2013; Rees et al., 2015). Proximity labeling relies on an enzyme, typically a biotin ligase, that is capable of covalently labeling, e.g. biotinylation, proteins in the immediate vicinity. Schmidtmann and coworkers combined this principle with CRISPR to develop the hybrid CasID approach, allowing biotinylation of proteins in the neighborhood of a specific genomic locus (figure 16A) (Schmidtmann et al., 2016). They fused a modified BirA (biotin ligase, BirA*) (Roux et al., 2012) to a dCas9, targeting telomeres and major or minor satellite sequences in mouse myoblast cells. Upon external addition of biotin, proteins located at the target locus are being biotinylated *in vivo*, allowing subsequent stringent affinity purification with streptavidin. CasID led to the isolation of different telomere associated proteins, including components of the shelterin complex, and known and unknown interacting proteins of major or minor satellite repeat sequences. The major difference with the above mentioned *in situ* gene-centered methods is the absence of *in vivo* cross-linking, since interactions do not need to be retained during the subsequent affinity purification. In principal every specific protein located at the genomic locus of interest is labeled with biotin, and can be isolated with high affinity using streptavidin.

In addition, biotinylation is performed in a time span of 24 hours, allowing accumulation of biotin labeled interactors. Even proteins that have a dynamic association with a specific locus, e.g. hit-and-run TFs (Doidy et al., 2016), which are otherwise difficult to pick up with gene-centered methods, will be labeled and pulled-down. However, the elaborated biotinylation time span also increases the labeling of non-interacting proteins which are just passing by the labeling radius, and together with the presence of endogenous biotinylated proteins, CasID is very prone to the pull-down of false positives. Therefore, quantitative MS analyses, relative to a suitable negative control, are necessary to identify the true interacting partners of a specific locus with CasID.

To diminish the amount of false positives, an alternative proximity-based gene-centered method has been developed, making use of an ascorbate peroxidase (APEX2) enzyme, which has a smaller labeling radius and shorter reaction time then BirA* (Myers et al., 2017). Similar to CasID, APEX2 is fused to dCas9, resulting in the CASPEX protein, and co-expressed with a locus specific gRNA. While CasID only needs external addition of biotin, APEX2 only biotinylated its neighborhood when biotin-phenol and hydrogen peroxide are available (figure 16B). Subsequent oxidation of biotin-phenol will result in phenoxyl radicals that react with surface exposed tyrosine residues on neighboring proteins. To avoid long-term presence and heaps of CASPEX in the cell, CASPEX expression is under control of an inducible promoter and will lead to a decrease in off-target binding and non-specific labeling of proteins. After induction, biotin-phenol is added to the cells for half an hour, followed by one minute of hydrogen peroxide treatment. Within this minute, proteins in close proximity of CASPEX are biotinylated. This strategy, called genomic locus proteomics (GLoPro), revealed known and unknown protein interactors of the single locus gene *hTERT*, and the *c-MYC* promoter in human cells (Myers et al., 2018). In addition to the use of a negative control which does not express a gRNA, a tiling strategy was implemented to correct for non-specific interacting proteins. Several gRNAs were designed for every single locus, and individually analyzed with GloPro. In this way, bona-fide interactors could be separated from co-purified proteins derived from off-target binding of the dCas9, by only selecting those MS signals that are retrieved with multiple gRNAs. As an extra advantage, this tiling strategy may also circumvent the loss of bona-fide interactions in case if Cas9 binding hinders protein association.

In comparison with CasID, CASPEX leads to a significant reduction of proximity labeling time and in combination with the tiling strategy it could result in a significant reduction of false positives. This advantage has resulted in the application of CASPEX in several studies to unravel the chromatin landscape of different genomic loci in human cells, including telomeres and centromeres (Myers et al., 2017; Gao et al.,

2018; Myers et al., 2018; Gao et al., 2019). In parallel to the development of GloPro, an alternative CRISPR-APEX2 gene-centered method was reported, called CAPLOCUS (figure 16C) (Qiu et al., 2019). Here, APEX2 is fused to the RNA binding protein MS2 and the locus specific gRNA is adapted with two MS2 RNA elements. CAPLOCUS was successfully used for the identification of interacting DNAs, RNAs and proteins for repetitive genomic regions (telomere, *C13*), as well as two single-copy loci (*C11* and *3'HS1*) in human cells (Qiu et al., 2019).

Although CRISPR-APEX2 applications are on the rise (figure 18), it should be taken into account that hydrogen peroxide delivery could lead to undesired cytotoxicity arising from oxidative stress signaling. A gene-centered tool that combines the simplicity and non-toxicity of CasID with the catalytic efficiency of APEX2 would be a more optimal approach. With the development of new promiscuous variants of BirA containing increased catalytic efficiency (Branon et al., 2018), these tools will be available in the future.

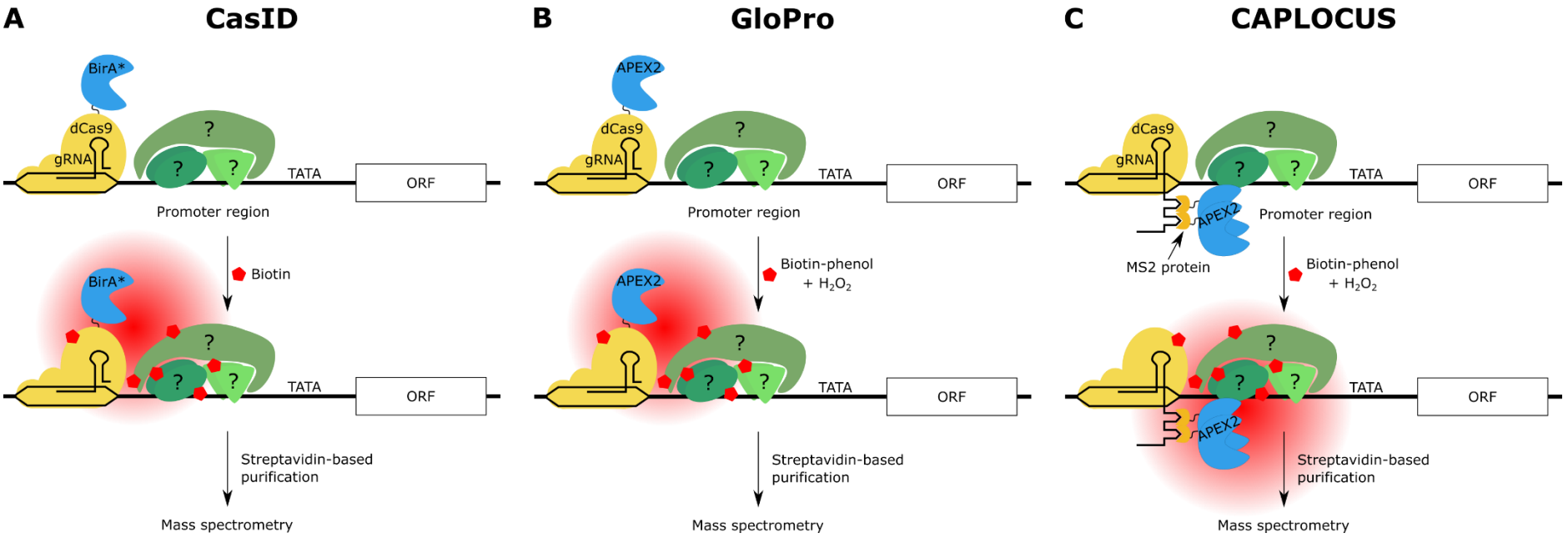


Figure 16: Scheme of CRISPR-dependent biotin-based proximity labeling approaches. **A)** Scheme for CasID. The BirA*-dCas9 fusion is directed to the desired target by sequence complementarity between gRNA and the promoter region. Upon addition of biotin to the medium, BirA* ligates biotin to lysine residues of proteins in close proximity. Biotinylated proteins can be pulled-down from the lysate with streptavidin and subjected to mass spectrometry. **B)** Scheme for GloPro. The dCas9-APEX2 fusion is directed to the target by sequence complementarity between gRNA and the promoter region. Upon addition of biotin-phenol and H_2O_2 to the medium, APEX2 will oxidate biotin-phenol to a phenoxyl radicals that react with surface exposed tyrosine residues on neighboring proteins. Biotinylated proteins can be pulled-down from the lysate with streptavidin and subjected to mass spectrometry. **C)** Scheme for CAPLOCUS. dCas9 is directed to the desired target by sequence complementarity between a gRNA, containing MS2 RNA elements, and the promoter region. The MS2-APEX2 fusion will associate with the modified gRNA and upon addition of biotin-phenol and H_2O_2 to the medium, APEX2 will oxidate biotin-phenol to a phenoxyl radicals that react with surface exposed tyrosine residues on neighboring proteins. Biotinylated proteins can be pulled-down from the lysate with streptavidin and subjected to mass spectrometry. ORF: open reading frame. TATA: TATA box.

Summary and perspectives

The fundamental process of gene transcription regulation determines which genes are expressed in which tissues depending on the cell type, developmental stage, environment, etc. Transcription regulation is the most important biological function of a cell and is primarily controlled by TFs and other regulatory proteins who interact with proximal and distal regulatory regions on the genome. Unraveling the identity and location of these transcription regulators is of utmost importance to understand an organism. As reviewed here, there are many different proteomic approaches available to detect and identify chromatin associated proteins and their complementary DNA recognition sequence. These techniques, subdivided in TF-centered and gene-centered ones, have been successfully used in many species.

EMSA and DNase I footprinting were the first applied TF-centered tools to identify protein-DNA interactions (Dahlberg et al., 1969; Fried and Crothers, 1981). But because of the limited number of DNA sequences that can be tested, the final consensus binding motifs have a low resolution. Also the low-throughput and laborious steps have led to the development of high-throughput TF-centered technologies like ChIP-seq. Although TF-centered techniques have been successfully used to determine protein-DNA interactions in many species, there is always need for prior knowledge of the protein of interest. There must be evidence that the protein is involved in the biological process of interest, a high quality ChIP-grade antibody must be available when performing ChIP-seq and the expression level must be high enough when *in situ* techniques are being applied. However, in many studies there is no knowledge about the proteins that are involved in a specific expression profile during a biological process. Here the transcriptional regulated genes are the baits of interest and need to be analyzed via gene-centered technique to identify the regulatory proteins. For many years protein microarrays and Y1H screens were the only options, but thanks to the development of mass spectrometry, new high-throughput techniques arose like DNA affinity chromatography and PICh.

Both groups (TF- and gene-centered) can be subdivided in *in vitro* and *in situ* techniques (figure 17). *In vitro* techniques like PBM and DNA affinity chromatography can deliver results relative fast with low amounts of input material. However, only interactions with high affinity are being retrieved because of the multiple washing steps that are applied. Because interactions are studied outside the native environment, obtained results should be interpreted with caution. False positive and negative results can be generated, because physiological properties are different and necessary PTM are absent in these systems. Also the binary nature (binding of one protein with one

DNA sequence) of most *in vitro* techniques leads to the non-detection of multimeric proteins. For these reasons *in vitro* techniques are more applied for a quick validation of the results obtained with *in situ* techniques. Although *in situ* techniques are more devious then *in vitro* techniques, more relevant data is generated. Thanks to the implementation of cross-linking, *in vivo* protein-DNA and protein-protein interactions are maintained during the different purification steps. In this way not only high affinity interactions are identified, but also transient and weak affinity interactions. And not only direct interactions between proteins and DNA are analyzed, also indirect interactions of coactivators and chromatin remodeling proteins can be studied as well as histone modifications (acetylation and methylation). However, *in vivo* approaches are more challenging because a DNA sequence may be present at a level of as few as one copy per cell. For TF-centered methods, increasing the starting material or amplifying the bound DNA sequences will result in enough DNA yield for further analysis. For *in vivo* gene-centered methods, one can also increase the starting material, but this will generally result in an amount that is not workable. Therefore, successful results have mostly been obtained by studying multi-copy genes and motif repeats.

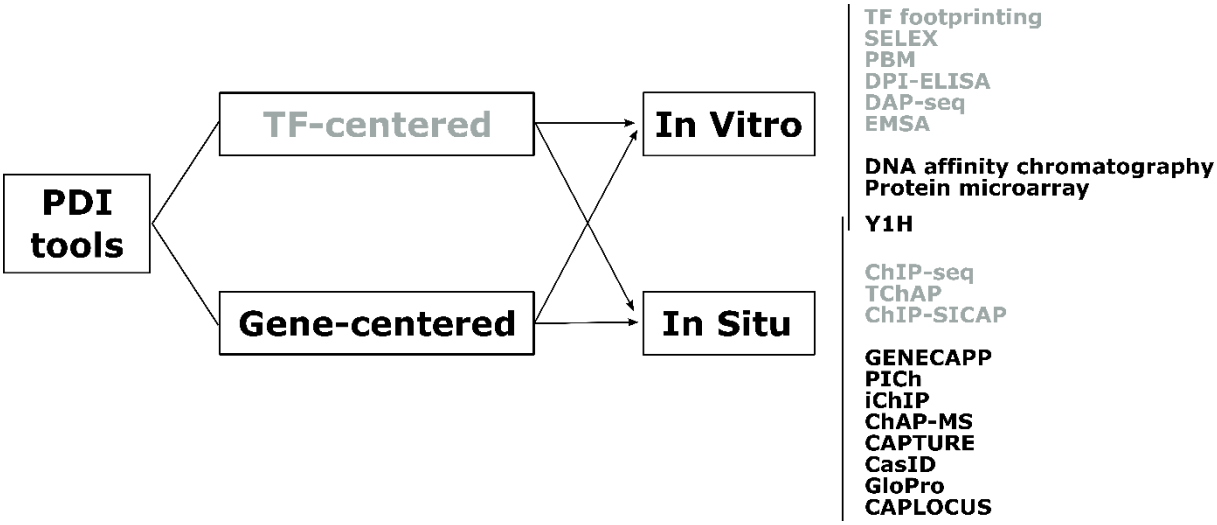


Figure 17: Overview of the different Protein-DNA interaction (PDI) tools.

Of all the above mentioned techniques, the gene-centered *in situ* methods are the most challenging and therefore the least well developed. As mentioned above, analyzing single copy genes is a hurdle because of the low yield obtained after affinity purification. Up to now, a handful single loci have been analyzed with a gene-centered technique in combination with quantitative MS. Although true interactors were identified, the low number of known activators and repressors in the final data indicate that the detection threshold is still a limiting factor. Future generations of more sensitive MS technologies could enable major progress in analyzing single copy genes.

Because one-step affinity purifications generally yield a big pool of specific and non-specific isolated proteins, a quantitative analysis is a key step in the identification of bona fide interactors, especially when single copy loci are studied. For the IRF-1 and Pax5 1A promoter, cells were compared containing transcriptional active or repressive forms of the promoter (Fujita et al., 2015). For the *GAL1* locus, a quantitative analysis was performed between cells containing the LexA DE and cells who did not (Byrum et al., 2012; Waldrip et al., 2014). Most of these quantitative MS analyses were obtained by SILAC, although progress has been made in the label free quantitative analysis (Waldrip et al., 2014; Myers et al., 2018).

Also fixation of the interactions is of great importance when performing a gene-centered analysis and should be optimized to be successful. Cross-linking conditions that are too harsh, will lead to inefficient reverse cross-linking and elution of proteins, while soft cross-linking condition will lead to the loss of interactors during the different extraction and affinity purification steps. Also the cross-linker that is used, should be carefully considered. Formaldehyde, which is generally used, is able to fixate not only PDI's but also PPI's, resulting in the isolation of big protein-DNA complexes and the co-isolation of a bunch of general chromatin related proteins during gene-centered analyses. This makes the subsequent MS analyses challenging and less efficient. An alternative cross-linker that solely fixates PDI's and is reversible could be a solution. However, not many fixators have these characteristics, except for cis-diammine dichloro platinum II (cis-DDP or cisplatin) which has been used to a limited extent for ChIP analyses (Chichiarelli et al., 2002; Cervoni et al., 2003; Chichiarelli et al., 2007). Cisplatin only forms protein-DNA complexes with a low reactivity towards histones (Pinto and Lippard, 1985) and the cross-linking can be reversed by the use of thiourea. Alternative cross-linking with cisplatin for gene-centered analysis of PDI's could lead to a more efficient identification of bona fide direct interactors with MS. However, proximity labeling gene-centered tools, like CasID and GloPro, avoid cross-linking and the corresponding pull-down of non-specific interactions.

With the discovery of CRISPR, new and rapidly developing approaches for gene-centered analysis are rising and will likely be the method of choice in the future (figure 18). Genetic engineering by inserting transgenes and making artificial repeats of the sequence of interest are being avoided by using CRISPR. Native loci can be targeted without changing the physiological chromatin structure. Although CRISPR is known for off-target binding, a tiling strategy, combining multiple gRNAs targeting one locus of interest, could decrease false positive hits (Myers et al., 2018). Successful gene centered experiments have been performed with CRISPR (Fujita and Fujii, 2013; Waldrip et al., 2014) and new variants are being developed which are making use of

the high affinity between biotin and streptavidin. One can fuse dCas9 to biotin and apply a stringent purification (Liu et al., 2017; Liu et al., 2018) or one can fuse dCas9 to a promiscuous biotin ligase (BirA*) or ascorbate peroxidase (APEX2), allowing for proximity-dependent biotin identification (Schmidtmann et al., 2016; Myers et al., 2017).

The list of tools for identification of protein-DNA interactions keeps expanding due to the discovery of engineered DNA-binding molecules and the development of more sensitive MS technologies. Although many techniques can be applied for many organisms, there is still one major gap for plants. Gene-centered analysis of a plant locus can only be obtained by performing a Y1H screen, outside the native environment and with a higher chance of false positive and negative results. In addition, Y1H screens only identify direct interactors, missing out the protein complexes generally present on chromatin. To obtain a genome-wide view of the protein complexes bound on specific plant genomic regions, gene-centered methods like PICh, ChAP-MS and iChIP are needed. Although, PICh has been applied once in barley for the analysis of centromeric chromatin (Zeng and Jiang, 2016), no other studies have reported the development or application of *in planta* gene-centered methods. There is need for the development of a new, plant specific method that can yield enough material for the identification of protein-DNA interactions via a quantitative MS analysis. Such a technique would be an important new platform for the elucidation of gene regulatory networks in plants.

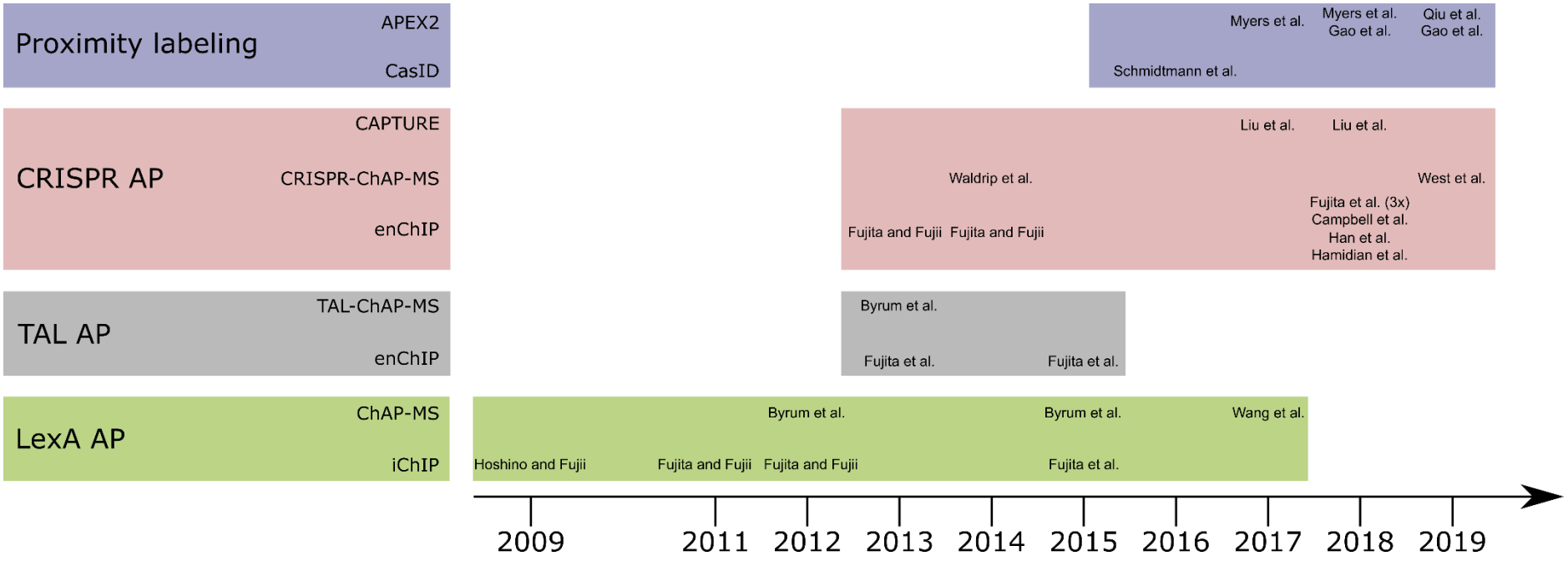


Figure 18: Time–line and overview of publication covering in situ gene-centered protein-DNA interaction methods. AP: affinity purification

References

Andrilenas KK, Penvose A, Siggers T (2015) Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes. Brief Funct Genomics **14:** 17-29

Antao JM, Mason JM, Dejardin J, Kingston RE (2012) Protein landscape at Drosophila melanogaster telomere-associated sequence repeats. Mol Cell Biol **32:** 2170-2182

Bajic M, Maher KA, Deal RB (2018) Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. Methods Mol Biol **1675:** 183-201

Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc **12:** 1659-1672

Bartocci C, Diedrich JK, Ouzounov I, Li J, Piunti A, Pasini D, Yates JR, 3rd, Lazzerini Denchi E (2014) Isolation of chromatin from dysfunctional telomeres reveals an important role for Ring1b in NHEJ-mediated chromosome fusions. Cell Rep **7:** 1320-1332

Brand LH, Henneges C, Schussler A, Kolukisaoglu HU, Koch G, Wallmeroth N, Hecker A, Thurow K, Zell A, Harter K, Wanke D (2013) Screening for protein-DNA interactions by automatable DNA-protein interaction ELISA. PLoS One **8:** e75177

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Brent R, Ptashne M (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43:** 729-736

Brown K, Takawira LT, O'Neill MM, Mizrachi E, Myburg AA, Hussey SG (2019) Identification and functional evaluation of accessible chromatin associated with wood formation in Eucalyptus grandis. New Phytol **223:** 1937-1951

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods **10:** 1213-1218

Burdo B, Gray J, Goetting-Minesky MP, Wittler B, Hunt M, Li T, Velliquette D, Thomas J, Gentzel I, dos Santos Brito M, Mejia-Guerra MK, Connolly LN, Qaisi D, Li W, Casas MI, Doseff AI, Grotewold E (2014) The Maize TFome--development of a transcription factor open reading frame collection for functional genomics. Plant J **80:** 356-366

Byrum SD, Raman A, Taverna SD, Tackett AJ (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep **2:** 198-205

Byrum SD, Taverna SD, Tackett AJ (2013) Purification of a specific native genomic locus for proteomic analysis. Nucleic Acids Res **41:** e195

Campbell AE, Shadle SC, Jagannathan S, Lim JW, Resnick R, Tawil R, van der Maarel SM, Tapscott SJ (2018) NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife **7**

Cao Y, Zeng H, Ku L, Ren Z, Han Y, Su H, Dou D, Liu H, Dong Y, Zhu F, Li T, Zhao Q, Chen Y (2020) ZmIBH1-1 regulates plant architecture in maize. J Exp Bot

Carlberg C, Molnár F (2014) The Basal Transcriptional Machinery. *In* Mechanisms of Gene Regulation. Springer, Dordrecht

Castrillo G, Turck F, Leveugle M, Lecharny A, Carbonero P, Coupland G, Paz-Ares J, Onate-Sanchez L (2011) Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of Arabidopsis transcription factors. PLoS One **6:** e21524

Cervoni L, Pietrangeli P, Chichiarelli S, Altieri F, Egistelli L, Turano C, Lascu I, Giartosio A (2003) In vivo cross-linking of nm23/nucleoside diphosphate kinase to the PDGF-A gene promoter. Mol Biol Rep **30:** 33-40

Chen CS, Zhu H (2006) Protein microarrays. Biotechniques **40:** 423, 425, 427 passim

Chichiarelli S, Coppari S, Turano C, Eufemi M, Altieri F, Ferraro A (2002) Immunoprecipitation of DNA-protein complexes cross-linked by cis-diamminedichloroplatinum. Anal Biochem **302:** 224-229

Chichiarelli S, Ferraro A, Altieri F, Eufemi M, Coppari S, Grillo C, Arcangeli V, Turano C (2007) The stress protein ERp57/GRP58 binds specific DNA sequences in HeLa cells. J Cell Physiol **210:** 343-351

Collas P (2010) The current state of chromatin immunoprecipitation. Mol Biotechnol **45:** 87-100

Dahlberg AE, Dingman CW, Peacock AC (1969) Electrophoretic characterization of bacterial polyribosomes in agarose-acrylamide composite gels. J Mol Biol **41:** 139-147

Dejardin J, Kingston RE (2009) Purification of proteins associated with specific genomic Loci. Cell **136:** 175-186

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES (2011) FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. Proc Natl Acad Sci U S A **108:** 6680-6685

Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004) A gateway-compatible yeast one-hybrid system. Genome Res **14:** 2093-2101

Doidy J, Li Y, Neymotin B, Edwards MB, Varala K, Gresham D, Coruzzi GM (2016) "Hit-and-Run" transcription: de novo transcription initiated by a transient bZIP1 "hit" persists after the "run". BMC Genomics **17:** 92

Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res **9:** 6505-6525

Fujii H, Fujita T (2015) Isolation of Specific Genomic Regions and Identification of Their Associated Molecules by Engineered DNA-Binding Molecule-Mediated Chromatin Immunoprecipitation (enChIP) Using the CRISPR System and TAL Proteins. Int J Mol Sci **16:** 21802-21812

Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, Fujii H (2013) Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci Rep **3:** 3171

Fujita T, Fujii H (2011) Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS One **6:** e26109

Fujita T, Fujii H (2012) Efficient isolation of specific genomic regions by insertional chromatin immunoprecipitation (iChIP) with a second-generation tagged LexA DNA-binding domain. Advances in Bioscience and Biotechnology **3:** 626-629

Fujita T, Fujii H (2013) Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun

Fujita T, Fujii H (2014) Identification of proteins associated with an IFNgamma-responsive promoter by a retroviral expression system for enChIP using CRISPR. PLoS One **9:** e103084

Fujita T, Kitaura F, Fujii H (2015) A critical role of the Thy28-MYH9 axis in B cell-specific expression of the Pax5 gene in chicken B cells. PLoS One **10:** e0116579

Fujita T, Kitaura F, Oji A, Tanigawa N, Yuno M, Ikawa M, Taniuchi I, Fujii H (2018) Transgenic mouse lines expressing the 3xFLAG-dCas9 protein for enChIP analysis. Genes Cells **23:** 318-325

Fujita T, Yuno M, Fujii H (2018) An enChIP system for the analysis of bacterial genome functions. BMC Res Notes **11:** 387

Fujita T, Yuno M, Fujii H (2018) enChIP systems using different CRISPR orthologues and epitope tags. BMC Res Notes **11:** 154

Fujita T, Yuno M, Okuzaki D, Ohki R, Fujii H (2015) Identification of non-coding RNAs associated with telomeres using a combination of enChIP and RNA sequencing. PLoS One **10:** e0123387

Gadgil H, Jurado LA, Jarrett HW (2001) DNA affinity chromatography of transcription factors. Anal Biochem **290:** 147-178

Galli M, Khakhar A, Lu Z, Chen Z, Sen S, Joshi T, Nemhauser JL, Schmitz RJ, Gallavotti A (2018) The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. Nat Commun **9:** 4526

Gao XD, Rodriguez TC, Sontheimer EJ (2019) Adapting dCas9-APEX2 for subnuclear proteomic profiling. Methods Enzymol **616:** 365-383

Gao XD, Tu LC, Mir A, Rodriguez T, Ding Y, Leszyk J, Dekker J, Shaffer SA, Zhu LJ, Wolfe SA, Sontheimer EJ (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods

Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeples M, Pu L, Liu Z, Breton G, Pruneda-Paz JL, Kim D, Kay SA, Walhout AJ, Ware D, Brady SM (2011) Enhanced Y1H assays for Arabidopsis. Nat Methods **8:** 1053-1055

Gijs M, Aerts A, Impens N, Baatout S, Luxen A (2016) Aptamers as radiopharmaceuticals for nuclear imaging and therapy. Nucl Med Biol **43:** 253-271

Godoy M, Franco-Zorrilla JM, Perez-Perez J, Oliveros JC, Lorenzo O, Solano R (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. Plant J **66:** 700-711

Gong W, He K, Covington M, Dinesh-Kumar SP, Snyder M, Harmer SL, Zhu YX, Deng XW (2008) The development of protein microarrays and their applications in DNA-protein and protein-protein interaction analyses of Arabidopsis transcription factors. Mol Plant **1:** 27-41

Hamidian A, Vaapil M, von Stedingk K, Fujita T, Persson CU, Eriksson P, Veerla S, De Preter K, Speleman F, Fujii H, Pahlman S, Mohlin S (2018) Promoter-associated proteins of EPAS1 identified by enChIP-MS - A putative role of HDX as a negative regulator. Biochem Biophys Res Commun **499:** 291-298

Han B, Zhou B, Qu Y, Gao B, Xu Y, Chung S, Tanaka H, Yang W, Giuliano AE, Cui X (2018) FOXC1-induced non-canonical WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer. Oncogene **37:** 1399-1408

Harrison MM, Jenkins BV, O'Connor-Giles KM, Wildonger J (2014) A CRISPR view of development. Genes Dev **28:** 1859-1872

Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc **2:** 1849-1861

Hoshino A, Fujii H (2009) Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. J Biosci Bioeng **108:** 446-449

Hsieh TF, Fischer RL (2005) Biology of chromatin dynamics. Annu Rev Plant Biol **56:** 327-351

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML (2015) UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res **43:** D117-122

Ide S, Dejardin J (2015) End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. Nat Commun **6:** 6674

Inze D, De Veylder L (2006) Cell cycle regulation in plant development. Annu Rev Genet **40:** 77-105

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res **45:** D1040-D1045

Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc **5:** 457-472

Kwon C, Chung IK (2004) Interaction of an Arabidopsis RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. J Biol Chem **279:** 12812-12818

Lambert JP, Pawson T, Gingras AC (2012) Mapping physical interactions within chromatin by proteomic approaches. Proteomics **12:** 1609-1622

Lane D, Prentki P, Chandler M (1992) Use of gel retardation to analyze protein-nucleic acid interactions. Microbiol Rev **56:** 509-528

Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell **128:** 707-719

Li C, Qiao Z, Qi W, Wang Q, Yuan Y, Yang X, Tang Y, Mei B, Lv Y, Zhao H, Xiao H, Song R (2015) Genome-wide characterization of cis-acting DNA targets reveals the transcriptional regulatory framework of opaque2 in maize. Plant Cell **27:** 532-545

Li W, Lin YC, Li Q, Shi R, Lin CY, Chen H, Chuang L, Qu GZ, Sederoff RR, Chiang VL (2014) A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue. Nat Protoc **9:** 2180-2193

Liang X, Nazarenus TJ, Stone JM (2008) Identification of a consensus DNA-binding site for the Arabidopsis thaliana SBP domain transcription factor, AtSPL14, and binding kinetics by surface plasmon resonance. Biochemistry **47:** 3645-3653

Lin YC, Li W, Sun YH, Kumari S, Wei H, Li Q, Tunlaya-Anukit S, Sederoff RR, Chiang VL (2013) SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in Populus trichocarpa. Plant Cell **25:** 4324-4341

Liu X, Zhang Y, Chen Y, Li M, Shao Z, Zhang MQ, Xu J (2018) CAPTURE: In Situ Analysis of Chromatin Composition of Endogenous Genomic Loci by Biotinylated dCas9. Curr Protoc Mol Biol**:** e64

Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, Dickerson KE, Xie S, Hon GC, Xuan Z, Zhang MQ, Shao Z, Xu J (2017) In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell **170:** 1028-1043 e1019

Ma J, Ptashne M (1988) Converting a eukaryotic transcriptional inhibitor into an activator. Cell **55:** 443-446

Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, Queitsch C, Bailey-Serres J, Sinha N, Brady SM, Deal RB (2018) Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. Plant Cell **30:** 15-36

Matteau D, Rodrigue S (2015) Precise Identification of DNA-Binding Proteins Genomic Location by Exonuclease Coupled Chromatin Immunoprecipitation (ChIP-exo). Methods Mol Biol **1334:** 173-193

Meraz MA, White JM, Sheehan KC, Bach EA, Rodig SJ, Dighe AS, Kaplan DH, Riley JK, Greenlund AC, Campbell D, Carver-Moore K, DuBois RN, Clark R, Aguet M, Schreiber RD (1996) Targeted disruption of the Stat1 gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway. Cell **84:** 431-442

Mitsuda N, Ikeda M, Takada S, Takiguchi Y, Kondou Y, Yoshizumi T, Fujita M, Shinozaki K, Matsui M, Ohme-Takagi M (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in Arabidopsis thaliana. Plant Cell Physiol **51:** 2145-2151

Mittler G, Butter F, Mann M (2009) A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res **19:** 284-293

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2009) In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res **16:** 353-369

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2013) TreeTFDB: an integrative database of the transcription factors from six economically important tree crops for functional predictions and comparative and functional genomics. DNA Res **20:** 151-162

Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat Methods

Myers SA, Wright J, Zhang F, Carr SA (2017) CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus. Molecular & Cellular Proteomics **16:** S63-S63

Nikolov DB, Burley SK (1997) RNA polymerase II transcription initiation: a structural view. Proc Natl Acad Sci U S A **94:** 15-22

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165:** 1280-1292

Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet **10:** 669-680

Pfluger J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. Curr Opin Plant Biol **10:** 645-652

Pinto AL, Lippard SJ (1985) Binding of the antitumor drug cis-diamminedichloroplatinum(II) (cisplatin) to DNA. Biochim Biophys Acta **780:** 167-180

Qiu W, Xu Z, Zhang M, Zhang D, Fan H, Li T, Wang Q, Liu P, Zhu Z, Du D, Tan M, Wen B, Liu Y (2019) Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. Nucleic Acids Res **47:** e52

Rafiee MR, Girardot C, Sigismondo G, Krijgsveld J (2016) Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. Mol Cell **64:** 624-635

Reece-Hoyes JS, Diallo A, Lajoie B, Kent A, Shrestha S, Kadreppa S, Pesyna C, Dekker J, Myers CL, Walhout AJ (2011) Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. Nat Methods **8:** 1059-1064

Reece-Hoyes JS, Marian Walhout AJ (2012) Yeast one-hybrid assays: a historical and technical perspective. Methods **57:** 441-447

Rees JS, Li XW, Perrett S, Lilley KS, Jackson AP (2015) Protein Neighbors and Proximity Proteomics. Mol Cell Proteomics **14:** 2848-2856

Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell **147:** 1408-1419

Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics **8:** 42

Richardt S, Lang D, Reski R, Frank W, Rensing SA (2007) PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. Plant Physiol **143:** 1452-1466

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science **290:** 2105-2110

Romeuf I, Tessier D, Dardevet M, Branlard G, Charmet G, Ravel C (2010) wDBTF: an integrated database resource for studying wheat transcription factor families. BMC Genomics **11:** 185

Rossi MJ, Lai WKM, Pugh BF (2018) Simplified ChIP-exo assays. Nat Commun **9:** 2842

Roux KJ (2013) Marked by association: techniques for proximity-dependent labeling of proteins in eukaryotic cells. Cell Mol Life Sci **70:** 3657-3664

Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol **196:** 801-810

Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H (2016) Determination of local chromatin composition by CasID. Nucleus **7:** 476-484

Schramm F, Larkindale J, Kiehlmann E, Ganguli A, Englich G, Vierling E, von Koskull-Doring P (2008) A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. Plant J **53:** 264-274

Schwechheimer C, Zourelidou M, Bevan MW (1998) Plant Transcription Factor Studies. Annu Rev Plant Physiol Plant Mol Biol **49:** 127-150

Shaikhali J, Noren L, de Dios Barajas-Lopez J, Srivastava V, Konig J, Sauer UH, Wingsle G, Dietz KJ, Strand A (2012) Redox-mediated mechanisms regulate DNA binding activity of the G-group of basic region leucine zipper (bZIP) transcription factors in Arabidopsis. J Biol Chem **287:** 27510-27525

Sims SH, Cha Y, Romine MF, Gao PQ, Gottlieb K, Deisseroth AB (1993) A novel interferon-inducible domain: structural and functional analysis of the human interferon regulatory factor 1 gene promoter. Mol Cell Biol **13:** 690-702

Smith LM, Shortreed MR, Olivier M (2011) To understand the whole, you must know the parts: unraveling the roles of protein-DNA interactions in genome regulation. Analyst **136:** 3060-3065

Sung MH, Guertin MJ, Baek S, Hager GL (2014) DNase footprint signatures are dictated by factor dynamics and DNA sequence. Mol Cell **56:** 275-285

Szutorisz H, Dillon N, Tora L (2005) The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci **30:** 593-599

Tao SC, Chen CS, Zhu H (2007) Applications of protein microarray technology. Comb Chem High Throughput Screen **10:** 706-718

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, Handakumbura PP, Xiong G, Wang C, Corwin J, Tsoukalas A, Zhang L, Ware D, Pauly M, Kliebenstein DJ, Dehesh K, Tagkopoulos I, Breton G, Pruneda-Paz JL, Ahnert SE, Kay SA, Hazen SP, Brady SM (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature **517:** 571-575

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249:** 505-510

van Werven FJ, Timmers HT (2006) The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. Nucleic Acids Res **34:** e33

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, Van de Peer Y, Inze D, De Veylder L (2005) Genome-wide identification of potential plant E2F target genes. Plant Physiol **139:** 316-328

Verkest A, Abeel T, Heyndrickx KS, Van Leene J, Lanz C, Van De Slijke E, De Winne N, Eeckhout D, Persiau G, Van Breusegem F, Inze D, Vandepoele K, De Jaeger G (2014) A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification. Plant Physiol **164:** 1122-1133

Vester B, Wengel J (2004) LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. Biochemistry **43:** 13233-13241

Viens A, Mechold U, Lehrmann H, Harel-Bellan A, Ogryzko V (2004) Use of protein biotinylation in vivo for chromatin immunoprecipitation. Anal Biochem **325:** 68-76

Viola IL, Uberti Manassero NG, Ripoll R, Gonzalez DH (2011) The Arabidopsis class I TCP transcription factor AtTCP11 is a developmental regulator with distinct DNA-binding properties due to the presence of a threonine residue at position 15 of the TCP domain. Biochem J **435:** 143-155

Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Wahls WP, Taverna SD, Raney KD, Tackett AJ (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics **9:** 1207-1211

Wang J, Lu J, Gu G, Liu Y (2011) In vitro DNA-binding profile of transcription factors: methods and new insights. J Endocrinol **210:** 15-27

Wang P, Byrum S, Fowler FC, Pal S, Tackett AJ, Tyler JK (2017) Proteomic identification of histone post-translational modifications and proteins enriched at a DNA double-strand break. Nucleic Acids Res **45:** 10923-10940

West KL, Byrum SD, Mackintosh SG, Edmondson RD, Taverna SD, Tackett AJ (2019) Proteomic characterization of the arsenic response locus in S. cerevisiae. Epigenetics **14:** 130-145

Wu CH, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Olivier M, Smith LM (2011) Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. PLoS One **6:** e26217

Xie Z, Hu S, Qian J, Blackshaw S, Zhu H (2011) Systematic characterization of protein-DNA interactions. Cell Mol Life Sci **68:** 1657-1668

Yanai K (2013) A modified yeast one-hybrid system for genome-wide identification of transcription factor binding sites. Methods Mol Biol **977:** 125-136

Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L, Grotewold E (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. Nucleic Acids Res **39:** D1118-1122

Zeng Z, Jiang J (2016) Isolation and Proteomics Analysis of Barley Centromeric Chromatin Using PICh. J Proteome Res **15:** 1875-1882

Zhang Q, Huang Q, Fang Q, Li H, Tang H, Zou G, Wang D, Li S, Bei W, Chen H, Li L, Zhou R (2020) Identification of genes regulated by the two-component system response regulator NarP of Actinobacillus pleuropneumoniae via DNA-affinity-purified sequencing. Microbiol Res **230:** 126343

Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J (2012) High-resolution mapping of open chromatin in the rice genome. Genome Res **22:** 151-162

Zhang W, Zhang T, Wu Y, Jiang J (2012) Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell **24:** 2719-2731

Zhu H, Snyder M (2003) Protein chip technology. Curr Opin Chem Biol **7:** 55-63

Zhu JY, Sun Y, Wang ZY (2012) Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq). Methods Mol Biol **876:** 173-188

Chapter 2

Exploring protein interaction landscapes.

Caroline Matthijs[1,2], Jelle Van Leene[1,2], Geert De Jaeger[1,2]

Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

Author contributions

CM performed literature study and wrote the chapter. JVL and GDJ supervised and complemented the writing.

Abstract

Proteins are essential laborers of a biological system and interact with each other, forming complexes and initiating signaling cascades for proper biological activity. Due to their different functionalities, dynamics and complexity, protein-protein interactions (PPIs) are a challenging field to investigate. Especially the identification of low affinity interactions and dynamic, transient interactions remain to this date a tough task to perform. Different tools exist unraveling binary and high-throughput protein-protein interactions. In this chapter, we give an overview of different *in situ* PPI techniques along with advantages, challenges and some successful applications in plants. We also discuss in more detail the uprising new tools for identification of weak transient interactions, including an in-house developed AP-MS tool used to reveal substrates of the plant TOR kinase.

Introduction

Proteins are the workhorses of a biological system and operate in such a way that an organism can function properly. They control and execute all cellular processes ranging from transcription and translation to sensing and responding to different stimuli. Executing this range of cellular functions requires collaboration and organization between different proteins. Proteins interact with each other, forming complexes and initiating signaling cascades for proper biological activity. This cellular network of protein-protein interactions (PPIs) only becomes more complex when we also consider the dynamic property of a protein. Expression and translation of proteins is tightly regulated and dependent on the developmental stage of a cell, the cell type and the environmental conditions. Often proteins are modified after translation (post translational modifications, PTMs) which influences the activity of a protein. In addition, controlled protein degradation, proteolysis, allows for removal of misfolded and non-functional proteins as well as maintaining homeostasis by controlling the abundancy of regulatory proteins.

To carry out a cellular function, proteins apply different mode of actions upon interaction with a partner. Interaction can lead to activation or inhibition of a protein due to conformational changes, whether or not accompanied by PTMs such as phosphorylation and ubiquitination. For example, during cell cycle interaction of a cyclin with a cyclin dependent kinase (CDK) results in the relocation of the PSTAIRE helix (or T-loop) away from the catalytic CDK site (Jeffrey et al., 1995). This results in partial activation of CDK, allowing ATP binding at the active site. To have a fully active cyclin-CDK complex that is able to phosphorylate target proteins, phosphorylation of the threonine residue 160 in the T-loop is needed by the CDK-activating kinase (CAK) (De Bondt et al., 1993; Draetta, 1997). On the other hand, inhibition of the active cyclin-CDK complex can be executed by interaction with a CDK inhibitor who blocks the catalytic CDK site (Elledge and Harper, 1994), by phosphorylation of tyrosine 15 and threonine 14 on CDK by WEE1 (Berry and Gould, 1996), or the ubiquitination of cyclin by the anaphase promoting complex (APC) and subsequent 26S proteasomal degradation (Glotzer et al., 1991; Kaspar et al., 2001). Another mode of action is transportation, which is well characterized in the kinesin, dynein and myosin families. This motor-cargo interaction allows the movement of proteins and organelles along microtubule and actin filaments (Karcher et al., 2002).

Multiple proteins can interact with each other to form a functional macromolecular complex. The largest multiprotein complex characterized to date is the mitochondrial complex I containing 45 subunits with a total mass of around 1 megadalton (MDa) (Fiedorczuk et al., 2016). Mutations in one of the subunits of a multiprotein complex is

often enough to inactivate the complex. For example, mutation studies of different APC subunits in *Arabidopsis* resulted in inactivation of the E3 ligase, the accumulation of cyclin substrates and mitotic arrest (Eloy et al., 2011; Wang et al., 2012; Wang et al., 2013; Eloy et al., 2015). The interactions between subunits of a protein complex are called stable interactions. Transient interactions on the other hand are temporary interactions, occurring under specific conditions or at a developmental phase. Most of the time these transient interactions monitor cellular processes. The APC in *Arabidopsis* is made up of 14 stably interacting subunits, but its activity is dependent on the transient interaction with co-activators (Eloy et al., 2015). Dependent on the cell cycle phase different co-activators associate with the APC. For example, at onset of mitosis the co-activator cell division cycle 20 (CDC20) activates the APC and provides substrate specificity, while at later stages of mitosis cell cycle switch 52 (CCS52, plant ortholog of human CDC20 homolog 1 (CDH1)) associates with the APC, resulting in targeting of CDC20 for proteasomal degradation and finalizing mitosis (Kramer et al., 2000; Yamano, 2019). While the transient interaction of co-activators with the APC is strong, also weak transient interactions are formed with substrates. For example, the well-known APC targets, cyclins, are bound by the co-activator and APC10 at specific time points during mitosis and are poly-ubiquitinated by the E3-ligase complex (Geley et al., 2001; Zhou et al., 2016). They rapidly dissociate from the complex and are subsequent targets for the 26S proteasome. So while cyclin forms a stable interaction with CDK during cell cycle, interaction with the APC is transient and rather weak.

It is clear that protein interactions are essential in a living organism, but their different functionalities, dynamics and complexity, makes it a challenging field to investigate. Techniques used for identification of PPIs can be divided in two groups, the binary techniques and the affinity-based methods in combination with mass spectrometry (MS). While binary techniques allow for identification of direct interactions between a bait protein and target (prey) protein, affinity-based methods are more recommended for the analysis of complex interaction landscapes including direct and indirect interactions. In this chapter an overview of different *in situ* PPI analysis tools is given along with advantages, challenges and some successful applications in plants. We also discuss in more detail the uprising new tools for identification of weak transient interactions, including an in-house developed AP-MS tool used to reveal substrates of the plant TOR kinase.

In situ binary techniques

The list of binary PPI techniques is elaborate, containing *in vitro*, *in situ* as well as *in silico* methods. Not all binary tools are discussed here, but several reviews can be found that elucidate the use of them for PPI research (Shoemaker and Panchenko, 2007, 2007; Stynen et al., 2012; Rao et al., 2014). In addition, several databases can be consulted to rummage in the binary protein interactome of numerous organisms (Chatr-Aryamontri et al., 2013; Orchard et al., 2014; Alonso-Lopez et al., 2019; Szklarczyk et al., 2019). Here we will focus on the most popular binary techniques and their application for plant PPI analysis.

Yeast two-hybrid

The most widely used binary method to investigate PPIs is a yeast two-hybrid (Y2H) assay (Fields and Song, 1989). This method makes use of the separable DNA-binding and transcriptional activation domain of a transcription factor. The DNA-binding domain (DBD) is fused to a bait protein, of which one wants to identify the interacting partners. The transcription activation domain (AD) is fused to a prey protein. Both fusion proteins are expressed in a yeast cell containing a reporter gene under control of upstream elements to which the DBD binds. If bait and prey protein interact with each other the DBD and AD will reconstitute an active transcription factor resulting in the expression of the reporter gene (figure 1A). In the classical approach the GAL4 DBD and AD are used in combination with the *LacZ* reporter (Fields and Song, 1989). Interaction leads to the detection of β-galactosidase activity. Over the years, many variations have been implemented to improve the technique. Yeast strains have been developed with mutations in different amino acid biosynthesis genes (*trp1, leu2, his3* and *ura3*) allowing selection for correct transformation events by using plasmids carrying genes that complement these mutations (Causier and Davies, 2002). Next to GAL4, the bacterial repressor protein LexA DBD in combination with the Eschericia coli B42 AD is frequently applied (Gyuris et al., 1993; Causier, 2004). To increase the stringency of the assay, yeast strains have been developed containing different reporter genes. For example, many Y2H screens with GAL4 are performed with the *LacZ* and *HIS3* reporter genes. Addition of the HIS3 inhibitor 3-amino-1,2,4-triazole (3-AT), at millimolar concentrations, reduces basal *HIS3* expression and therefore also false positive results. Other common reporters are *LEU2*, *URA3*, *LYS2*, *ADE2*, *gusA*, *GFP* and *MEL1* (Causier, 2004). An advantage is that not only interactions between known proteins can be analyzed with Y2H, but one can also screen cDNA or open reading frame (ORF) libraries to find unknown interaction partners of a specific bait protein or unravel a whole new PPI network. This high-throughput screening allows to investigate the protein interactome of a specific cell type or developmental stage. The use of cDNA

libraries ensures that Y2H is also applicable to organisms with a non-annotated genome. Different *Arabidopsis* cDNA and ORF libraries are available for GAL4 and LexA based Y2H screens (Burkle et al., 2005; Mitsuda et al., 2010; Braun et al., 2011; Erffelinck et al., 2018; Matiolli and Melotto, 2018). The most extensive proteome-wide binary protein-protein interaction map for *Arabidopsis* is generated with Y2H, revealing 6,200 interactions between 2,700 proteins (Braun et al., 2011). Recent years, Y2H screens are combined with next generation sequencing (NGS) (Yu et al., 2011; Trigg et al., 2017; Erffelinck et al., 2018). Erffelinck and co-workers designed an inexpensive and easy to apply Y2H system combining a cDNA library screen with a pool-based NGS-strategy. This allowed them to identify known and unknown interactions of the jasmonate (JA) signaling cascade of *Arabidopsis* (Erffelinck et al., 2018).

Although Y2H is a popular technique for the identification of binary protein interactions, it still has some drawbacks. Y2H is prone to contamination by false positives. This can be reduced by implementing multiple reporter genes. However, auto-activation of the reporter gene by the bait protein remains a hurdle, and is a general problem when screening a cDNA library (Serebriiskii et al., 2000). Therefore, it is important to test a bait protein for auto-activation before the Y2H screen is executed. Analyzing auto-activating bait proteins with a nuclear two-hybrid system can still be done by implementing RNA polymerase III (RNA POL III) (Marsolier et al., 1997; Petrascheck et al., 2001) or using the repressed transactivator (RTA) system (Hirst et al., 2001). Both alternatives and others have been reviewed in detail by Stynen and co-workers (Stynen et al., 2012). Another drawback is missing out on true interactions (false negatives) because of different reasons. For instance, to identify an interaction, bait and prey protein need to localize to the nucleus. However, not every protein is able to do so. Furthermore, proteins that repress gene expression will lead to false negative interpretation. If one wants to study non-yeast related interactions, e.g. *Arabidopsis* proteins, analyses are performed outside the native cellular context leading to false negatives and false positives, causing the need for *in planta* validation. Some interactions depend on PTMs, mechanisms that may be absent in yeast, such as tyrosine phosphorylation. Other non-yeast related proteins may be toxic for the yeast cell.

Because of its binary nature, only direct interactions are observed with Y2H. However, adaptations to the Y2H system have resulted in the development of yeast three-hybrid systems (Y3H) (Stynen et al., 2012). Next to the bait and prey protein, a third component is added, which may be a posttranslational modifier crucial for PTM of the bait or prey protein prior to interaction, or a third protein as a bridging molecule is included to identify indirect interactors.

To allow two-hybrid analysis in a native plant context, the *Arabidopsis* protoplast two-hybrid (P2H) system was developed (Ehlert et al., 2006). P2H uses the GAL4 system to activate a GUS reporter upon PPI. Gateway compatible vectors were generated, enabling high-throughput screening. To correct for auto-activation/repression, a control construct was developed allowing expression of a HA-tagged bait and/or prey protein. To correct for variation in transfection efficiency a second reporter construct was implemented, containing the synthetic neuraminidase reporter. Ehlert and co-workers could demonstrate the applicability of this system for the analysis of bZIP hetero-dimerization (Ehlert et al., 2006). In addition, confirmation of interaction between the RING-type copine McCPN1 (COPINE1) from a halophyte ice plant (Mesembryanthemum crystallinum L.) and argonaute 4 (AGO4) was obtained by P2H analysis (Li et al., 2014).

The split-ubiquitin system

Another yeast-based method that allows identification of binary protein interactions regardless of their localization, is the split-ubiquitin system (Johnsson and Varshavsky, 1994). Here, the C-terminal and N-terminal part of ubiquitin (Cub and Nub), a highly conserved 76-amino-acid regulatory protein, are split and each part is fused to a bait or prey protein (figure 1C). To circumvent spontaneous reassociation of both ubiquitin parts, Nub is mutated (I13G). Originally Cub is fused to a mouse dihydrofolate reductase and Human influenza hemagglutinin (mDHFR-HA). Upon interaction, a functional ubiquitin is formed and becomes a target of an ubiquitin-specific proteases (UBPs) that cleaves off mDHFR-HA. This results in a shift on western blot using anti-HA antibodies. However, to allow for library screening, an alternative reporter system with LexA-VP16 (herpes simplex virus (HSV)-encoded transcriptional activator protein) was developed which leads to activation of reporter genes (e.g. *LacZ* and *HIS3*) upon interaction (Stagljar et al., 1998). To avoid auto-activation, Cub-LexA-VP16 must be fused to a bait protein that is kept out of the nucleus and therefore the LexA-VP16 system is almost always performed with membrane-bound bait proteins (figure 1C). Alternatively, bait proteins can be fused at the N-terminus with the *S. cerevisiae* integral membrane protein Ost4 to direct the bait protein to the membrane (Mockli et al., 2007). This Ost4-based membrane localization allows for identification of interactions between transcription factors and co-regulators (transcription activators and repressors), and for example has been used to screen an *Arabidopsis* seedling cDNA library for interactions partners of the nuclei localized, RNA binding PNM1 protein (Hammani et al., 2011). In addition, an alternative split-ubiquitin system was established avoiding transcriptional read out (Laser et al., 2000). This alternative system makes use of the N-end rule (Dohmen et al., 1991; Varshavsky, 1996) which

states that proteins with N-terminal basic (e.g., arginine) or bulky hydrophobic amino acids are promoted for protein degradation in an ubiquitin-dependent manner. Replacing LexA-VP16 with the reporter protein URA3, with a N-terminal arginine, results in cleavage and subsequent degradation of URA3. URA3 removal leads to survival of the yeast cells on 5-FOA, a substrate for the production of the toxic compound 5-fluorouracil by URA3 (Laser et al., 2000).

An example of applying split-ubiquitin to explore plant PPIs, is the investigation of the membrane-linked interactome of *Arabidopsis* (Jones et al., 2014). Jones and co-workers screened a cDNA library of 3286 membrane and signaling proteins from *Arabidopsis* (Lalonde et al., 2010). They identified 12,102 interactions between 1523 proteins, most of them (>99%) being unknown interactions. Validation of a subset of interactions was performed by an *in planta* split–green fluorescent protein (GFP) interaction assay (see further).

A plant split-ubiquitin system was developed in *Arabidopsis* and tobacco protoplasts, using AtUBQ11 and HA-Cub-GFP as reporter protein (Rahim et al., 2009). Upon interaction, GFP is cleaved off and subsequent western blot analysis with anti-HA antibodies results in a shift of the HA tagged bait protein. Interactions between components of the translocon complex (TOC) were visualized, however the system in plant protoplasts displayed a high amount of background cleavage (Rahim et al., 2009). It was postulated that because of the extensive quantity of plant UBPs, the overall activity of UBPs in plants is higher than in yeast, making the plant split-ubiquitin system a suboptimal approach for PPI analysis.

The split-luciferase system

Another protein fragment complementation assay is the split-luciferase system, which uses bioluminescence to visualize PPIs (Ozawa et al., 2001). Luciferase is a protein that oxidates the membrane-permeating substrate luciferin, accompanied by emitted light. Similar to the split-ubiquitin system, luciferase is separated in a C- and N-terminal part, both fused to a bait or prey protein. Upon interaction, an active luciferase is formed and the emitted light is measured by a luminometer (figure 1B). Because luminescence signal is dependent on the transformation efficiency, signals are normalized to a luciferase of different origin (e.g. *Renilla reniformis* luciferase versus *Pyrophorus plagiophthalamus* luciferase) (Kato et al., 2010). This system is not dependent on the read out of transcriptional activation, allowing analysis of cytosolic, membrane associated and transcription related proteins. Furthermore, the system is reversible and permits the visualization of dynamic PPIs in near-real-time manner (Stefan et al., 2007). The system is also compatible with stimuli treatment to induce or

disrupt interactions (Li et al., 2011). Improvements have been made by using different luciferase proteins with much stronger brightness or which emit different colors of light (Villalobos et al., 2008; Stynen et al., 2012). Although some studies have stated that the intensity of luminescence is correlated with the affinity strength of the PPI, Dale et al. showed that this is not the case and that the split-luciferase system has to be seen as a qualitative analysis (Dale et al., 2016).

A split-luciferase system in *Arabidopsis* protoplast was developed by Fujikawa and Kato, revealing the interactions between nuclear proteins and membrane proteins (Fujikawa and Kato, 2007). Li et al. used the split-luciferase system for a large-scale PPI survey between auxin response factors and Aux/IAA proteins in *Arabidopsis* mesophyll protoplasts (Li et al., 2011). Because luciferase-analogous are absent in plants, the background luminescence is relative low making this approach appropriate for analyzing weak transient interactions like ARF-Aux/IAA. Further on, split-luciferase assays have also been reported in tomato and rice protoplasts (Fujikawa et al., 2014; Singh et al., 2014).

Bimolecular fluorescence complementation

A disadvantage of the split-luciferase system is the generation of a diffused light signal, preventing visualization of the subcellular localization of PPIs. Bimolecular fluorescence complementation (BiFC) is another protein fragment complementation assay using fluorescence to reveal PPIs and their correlated subcellular localization in living cells (Magliery et al., 2005). Here a fluorescent protein is split in two, both parts are fused to a bait or prey protein and upon interaction the fluorescent protein is reconstituted and will emit light of a specific wavelength after excitation by an external light source (figure 1D). This light absorption-based excitation of fluorescent molecules results in a concentrated light signal, perfect for subcellular localization studies using a laser scanning confocal microscopy. Furthermore, interaction results in a strong stabilization of the fluorescent protein making it an ideal system to study weak transient interactions, however, this makes analysis of PPIs dynamics impossible. Another advantage is that no exogenous agents are required. Initially, a green fluorescent protein (GFP) was used (Magliery et al., 2005), but over the years several natural and genetically optimized fluorescent proteins became available, among which enhanced GFP (eGFP), yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), the YFP variant Venus, and the monomeric Kusabira-Green fluorescent mutant (mKG2) (Day and Davidson, 2009; Miller et al., 2015). Alternative systems have been developed using multiple FPs to detect different PPIs in parallel or to study competition between different binding partners of a bait protein (Hu and Kerppola, 2003; Kodama and Wada, 2009; Stynen et al., 2012). However, the use of fluorescent proteins is

accompanied with several disadvantages, including photobleaching and phototoxicity. The repeated cycles of excitation and emission, using illumination from lasers, can lead to damage and destruction of the fluorescent protein (photobleaching) or may become toxic for the cell. Autofluorescence should also be taken into account, especially for plants which contain different endogenous fluorophores like chlorophyll and lignin. Lastly, self-assembly of the fluorescent protein halves without a PPI event has been observed. This is often true for proteins that are at close proximity, with equal localization, but do not interact. To avoid these false positives, it is crucial to include appropriate internal controls like a mutated version of the bait protein, or an unrelated prey protein with the same cellular localization (Horstman et al., 2014).

BiFC is an easy system that has been applied in different organisms, ranging from mammalian cells to several plant species. Especially in plants, BiFC studies have become a popular method to analyze binary PPIs, and extensive overviews of in planta BiFC application have been reported (Bhat et al., 2006; Ohad et al., 2007; Ohad and Yalovsky, 2010; Miller et al., 2015). For example, a large-scale BiFC screen was performed to map PPIs between core cell cycle proteins of *Arabidopsis* (Boruc et al., 2010). A pair-wise analysis between 58 cell cycle regulatory proteins was conducted, resulting in the identification of 341 PPIs in leaf epidermal cells of tobacco. Among these cell cycle related PPIs, 63 were already described in literature and an extra 33 unknown PPIs were confirmed by a Y2H screen. A high-throughput *Arabidopsis* cDNA-library BiFC screen in *Arabidopsis* protoplast revealed 8 interactor candidates of the Calcium Dependent Protein Kinase 3 (CPK3) (Berendzen et al., 2012). To allow for high-throughput screening, BiFC was combined with fluorescence assisted cell sorting (FACS). In short, *Arabidopsis* protoplast are transfected with a bait plasmid and a plasmid mix encoding *Arabidopsis* cDNA. By using a flow cytometer and FACS, fluorescent cells are collected and corresponding plasmid DNA is isolated and transformed in bacteria. A subsequent round of transfection, fluorescent cell collection and plasmid isolation is performed. In a final transfection event, plasmids of protoplast with positive BiFC signal are sequenced to identify the corresponding cDNA and associated encoded protein. This high-throughput BiFC screening method in *Arabidopsis* protoplast was time-consuming, taking 3 to 4 weeks, consequently, it is not an optimal approach for genome wide PPI screening.

Fluorescence/Bioluminescence resonance energy transfer

Another method that makes use of fluorescent proteins, but does not rely on the complementation of protein fragments, is fluorescence resonance energy transfer (FRET) (Periasamy et al., 2012; Bucherl et al., 2014). This approach makes use of two fluorescent proteins, one being the donor, the other the acceptor, both fused to a bait

or prey protein. Upon interaction, an external light source will excite the donor which will emit fluorescence/energy that is subsequently transferred to the acceptor, resulting in the emitted light of a specific wavelength (figure 1E). The most frequently used combination of fluorescent proteins is CFP as donor and YFP as acceptor (Day and Davidson, 2009). CFP is excited at a wavelength around 435 nm and will emit light of around 475 nm. This emitted light overlaps with the YFP absorption spectra and will result in the excitation of YFP and the emission of light around 527 nm. Transfer of energy between donor and acceptor will only occur when both are in close proximity (around 10-100 Å) and is accompanied by quenching of the donor emission. As an alternative, the donor can be substituted by luciferase (bioluminescence resonance energy transfer, BRET) avoiding the fluorescent excitation by illumination from lasers and the associated disadvantages of photobleaching, phototoxicity and autofluorescence (Pfleger and Eidne, 2006; Xie et al., 2011). With FRET/BRET it is not only possible to visualize the subcellular localization of the PPI, it also allows the analysis of PPI dynamics, PPI response on chemicals and stimuli, and FRET efficiency can be linked with PPI affinity (Coriano et al., 2016; Lin et al., 2018). An overview of FRET and BRET analyses in different plant species have been published (Bhat et al., 2006; Bucherl et al., 2014).

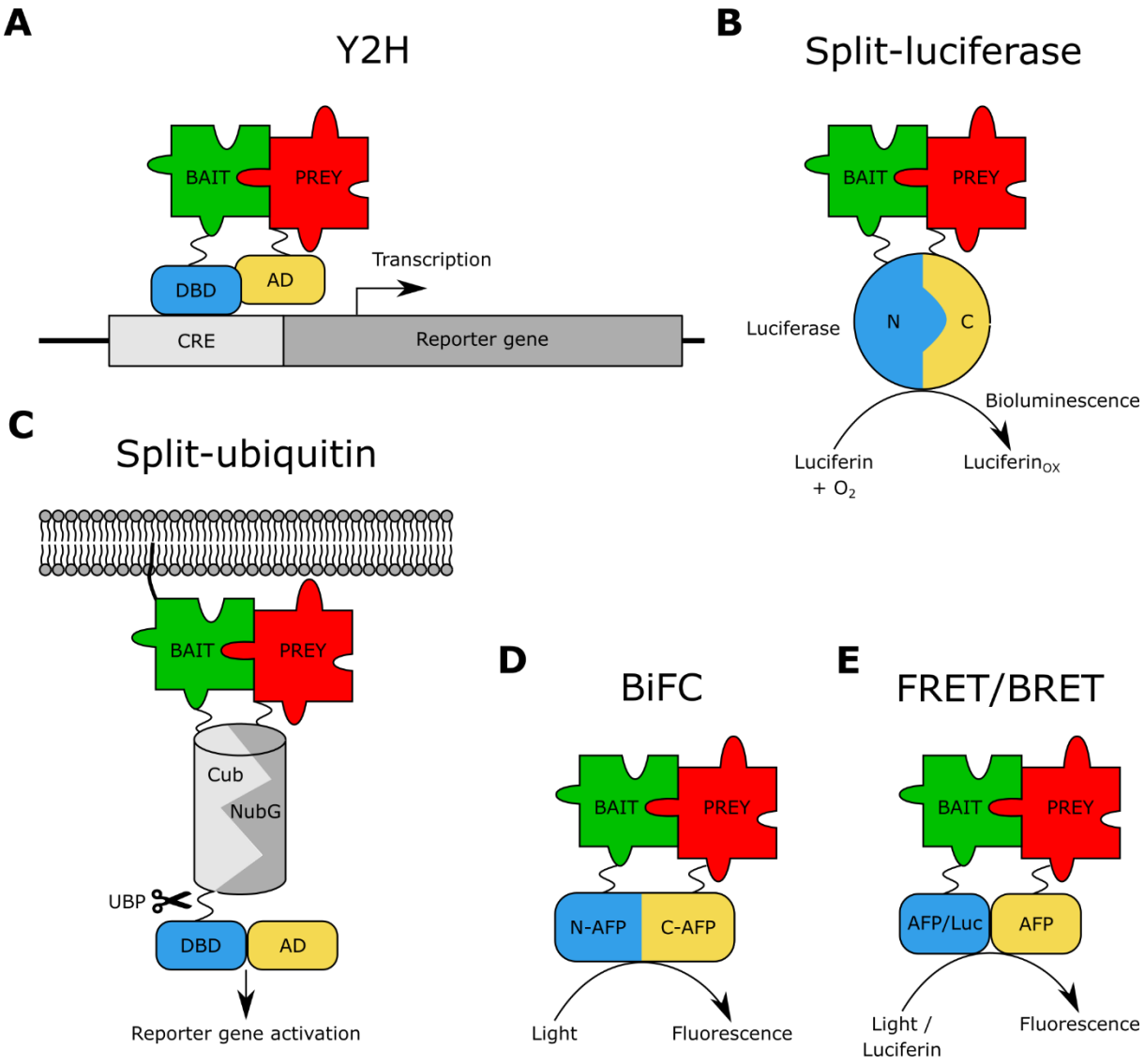


Figure 1: Schematic representation of different in situ binary protein-protein interaction techniques. **A)** The yeast two-hybrid (Y2H) system detects protein-protein interactions (PPIs) by activation of a reporter gene. Upon interaction between bait and prey protein a DNA binding domain (DBD) and activation domain (AD) are brought together and reconstitute an active transcription factor. CRE: cis regulatory element. **B)** The split-luciferase system detects PPIs via bioluminescence, derived from the reconstitution of the luciferase enzyme upon interaction between bait and prey protein. External added luciferin will be oxidized by luciferase, accompanied by emitted light. N: N-terminal part of luciferase. C: C-terminal part of luciferase. **C)** The split-ubiquitin system detects PPIs by activation of a reporter gene. Upon interaction between a membrane anchored bait protein and prey protein, ubiquitin is reconstituted, making it a target for an ubiquitin-specific proteases (UBP), which results in cleaving off a transcription activator. Cub: C-terminal part of ubiquitin. NubG: mutated N-terminal part of ubiquitin to circumvent spontaneous reassociation. **D)** Bimolecular fluorescence complementation (BiFC) detects PPIs via fluorescence, derived from the reconstitution of an auto fluorescent protein (AFP) upon interaction between bait and prey protein. **E)** Fluorescence/Bioluminescence resonance energy transfer (FRET/BRET) makes use of AFPs and/or luciferase (Luc), one being the donor, the other the acceptor, both fused to a bait or prey protein. Upon interaction, an external light source will excite the donor which will emit fluorescence/energy that is subsequently transferred to the acceptor, resulting in the emitted light of a specific wavelength.

Affinity-based methods in combination with mass spectrometry

As previously stated, multiple proteins can interact with each other to form a functional macromolecular complex, which in turn can interact with co-activators, regulators and substrates. To be able to elucidate the interaction landscape of multiprotein complexes, affinity-based methods have been developed and are more suited than binary PPI tools (Bontinck et al., 2018). These methods specifically purify a bait protein from a biological sample under near-physiological conditions, keeping PPIs intact, and are combined with mass spectrometry to identify the co-purified proteins.

Co-immunoprecipitation and single-step affinity purification

A method that make use of a bait-specific antibody to purify the endogenous target protein and its interactors is called immunoprecipitation (IP, figure 2A) (Masters, 2004). Initially, IP was applied to analyze the interaction between two proteins, isolating the bait with a specific antibody, and verifying the presence of a prey protein after IP on western blot with a prey-specific antibody. However, this co-IP strategy is dependent on the existence of bait- and prey-specific antibodies, which are very often not available. Therefore, an adjustment was made, fusing the bait and/or prey protein with different tags for which generic antibodies exist. This adjustment made co-IP a powerful tool and has become a standard method for identification and validation of PPIs *in vivo*. Tags that are commonly used are c-myc, HA, FLAG, and GFP, with GFP as the most frequently used one in plants, allowing localization studies in parallel (Dedecker et al., 2015). As an example of co-IP in plants, interaction between TPLATE and TML was confirmed by fusing TPLATE to HA and TML to FLAG (Gadeyne et al., 2014). Anti-FLAG co-IP and subsequent western blot analysis revealed the presence of HA-tagged TPLATE in the final TML pull-down sample.

To allow for a large-scale detection of unknown protein interactions, co-IP can be combined with mass spectrometry, revealing unknown direct interactors, indirect interactions and the isolation of protein complexes. In this way, the *Arabidopsis* mediator complex was purified, using an antibody against the mediator subunit 6 (Med6) (Backstrom et al., 2007). In addition, the use of antibodies becomes unnecessary when affinity tags are implemented (affinity purification, AP) (figure 2A). Different affinity tags exist and are reviewed elsewhere (Lichty et al., 2005; Kimple et al., 2013). The advantage of using affinity tags is that they generally have a stronger affinity towards their binding partner than antibodies towards their epitope, which enables the development of generic purification methods. The superior strength allows for more stringent washing conditions during an AP protocol, resulting in a better signal-to-noise ratio. The binding between biotin and streptavidin represents one of the

strongest non-covalent interactions known (Kd = 10^{-15} M) and has been used to identify interactors of the TATA-box-binding protein (TBP) in rice (Zhong et al., 2003). Zhong and co-workers fused TBP to a biotin peptide and used magnetic streptavidin beads for affinity purification. To allow elution of TBP and its interacting partners, a tobacco etch virus (TEV) protease cleavage site was incorporated between TBP and the biotin peptide. In addition, TEV protease-based elution resulted in the reduction of endogenous biotinylated protein contaminants in the final eluate.

Affinity purification has been successfully performed in numerous plant studies (Mravec et al., 2011; Smaczniak et al., 2012; De Rybel et al., 2013; Debernardi et al., 2014). However, the major pitfall during single-step affinity purifications, is the non-specific isolation of abundant and promiscuous proteins, often representing more than 90% of the AP-MS identified proteins (Trinkle-Mulcahy et al., 2008). To correct for this, different strategies can be applied, like the above mentioned implementation of a protease cleavage site whether or not followed by a consecutive affinity step (see tandem affinity purification section). Another approach is the subtraction of a background list, generated by control experiments or by combining multiple purifications of unrelated bait proteins. In addition, a quantitative MS analysis can be implemented for identification of bona fide and weakly interacting proteins (Nesvizhskii, 2012).

Tandem affinity purification

To avoid the isolation of non-specific interacting proteins with AP-MS as much as possible, a double-step affinity purification (TAP) strategy was developed (figure 2A) (Li, 2011). The first TAP application was performed in yeast, using a TAP tag which contains a calmodulin binding protein (CBP) and two Protein A (ProtA), separated by a TEV protease cleavage site (Rigaut et al., 1999; Gerace and Moazed, 2015). Over time, different TAP tags have been generated, with the TAPi tag (figure 2B-1) (Rohila et al., 2004) and GS-tag (Van Leene et al., 2008) being the most widely adopted in plant studies (Dedecker et al., 2015). Further optimization of the GS-tag has resulted in a tag that is more suitable for studying unstable proteins with a high turnover, called the GS^{rhino}-tag (Van Leene et al., 2015). GS^{rhino} consists of two protein G tags and the streptavidin-binding peptide (SBP) separated by two rhinovirus 3C protease cleavage sites (figure 2B-2). In a first purification step, the bait protein and interacting partners are isolated through high-affinity binding on an IgG resin. After a first washing step, the rhinovirus 3C protease is added for a gentle and specific elution by recognizing and snipping the tandem cleavage sites. In a subsequent purification step, the bait protein complex is trapped through binding to streptavidin-conjugated beads. Residual proteases and contaminating proteins are removed in a second washing step. Elution

of a pure protein complex is obtained by addition of desthiobiotin, which will compete with the SBP-streptavidin interaction, leading to a gentle and easy elution.

TAP has been applied in a wide range of organisms, including yeast, insects, mammalian cells, bacteria and different plant species (Li, 2011; Dedecker et al., 2015). It has been frequently used for large-scale analyses to map specific protein landscapes. As an example, TAP has been applied in *Arabidopsis* cell suspension culture to unravel the cell cycle interactome (Van Leene et al., 2010). Analysis of 102 cell cycle related bait proteins resulted in the identification of 857 interactions among 393 proteins. Reverse TAP experiments, using co-purified proteins as bait, confirmed the initial interactions, and screening of public PPI databases revealed the presence of 150 known or predicted interactions. In addition, plant-based TAP experiments have revealed TF and E3-ligase complexes in *Arabidopsis*, rice and Medicago truncatula (Van Leene et al., 2011; Eloy et al., 2012; Dedecker et al., 2016; Goossens et al., 2016), and the dynamic chances of a chromatin remodeling complex in maize (Nelissen et al., 2015; Bontinck et al., 2018). Development of a multifunctional TAP-tag, combining the fluorescent protein YFP with SBP (GS^{yellow}, figure 2B-3), allowed not only the affinity purification of dicot and monocot protein complexes, but also their subcellular localization and, in case of chromatin related proteins, their DNA-binding landscape (Besbrugge et al., 2018).

Although TAP was originally developed to minimize the amount of non-specific interacting proteins, the increasing sensitivity of mass spectrometers still leads to a high amount of background proteins and still require careful background filtering to reduce false positives. Comprehensive background lists for *Arabidopsis* cell cultures, seedlings, rice cell cultures, maize and *Medicago* are publicly available (Van Leene et al., 2015; Dedecker et al., 2016; Goossens et al., 2016; Besbrugge et al., 2018), and in the recent years, TAP has been combined with label free semi-quantitative MS analysis (Meyer and Selbach, 2015; Van Leene et al., 2019). TAP-MS generally misses out on very low abundant proteins, proteins expressed only in rare cell types or developmental stages, and proteins with poor solubility like integral membrane proteins. TAP analyses are also prone to false negative results. Because of the iterative washing steps, needed to avoid a high amount of false positives, weak and transient interacting proteins are lost. This was clearly demonstrated when the plant TOR kinase PPI landscape was explored (Van Leene et al., 2019). TAP studies revealed the isolation of mainly stable interacting proteins of the TOR complex. However, an one-step pull-down protocol combined with label-free quantitative MS was more efficient for the discovery of putative TOR substrates, as an increased

overlap with a TOR dependent phosphoproteome was found (see further for more detail).

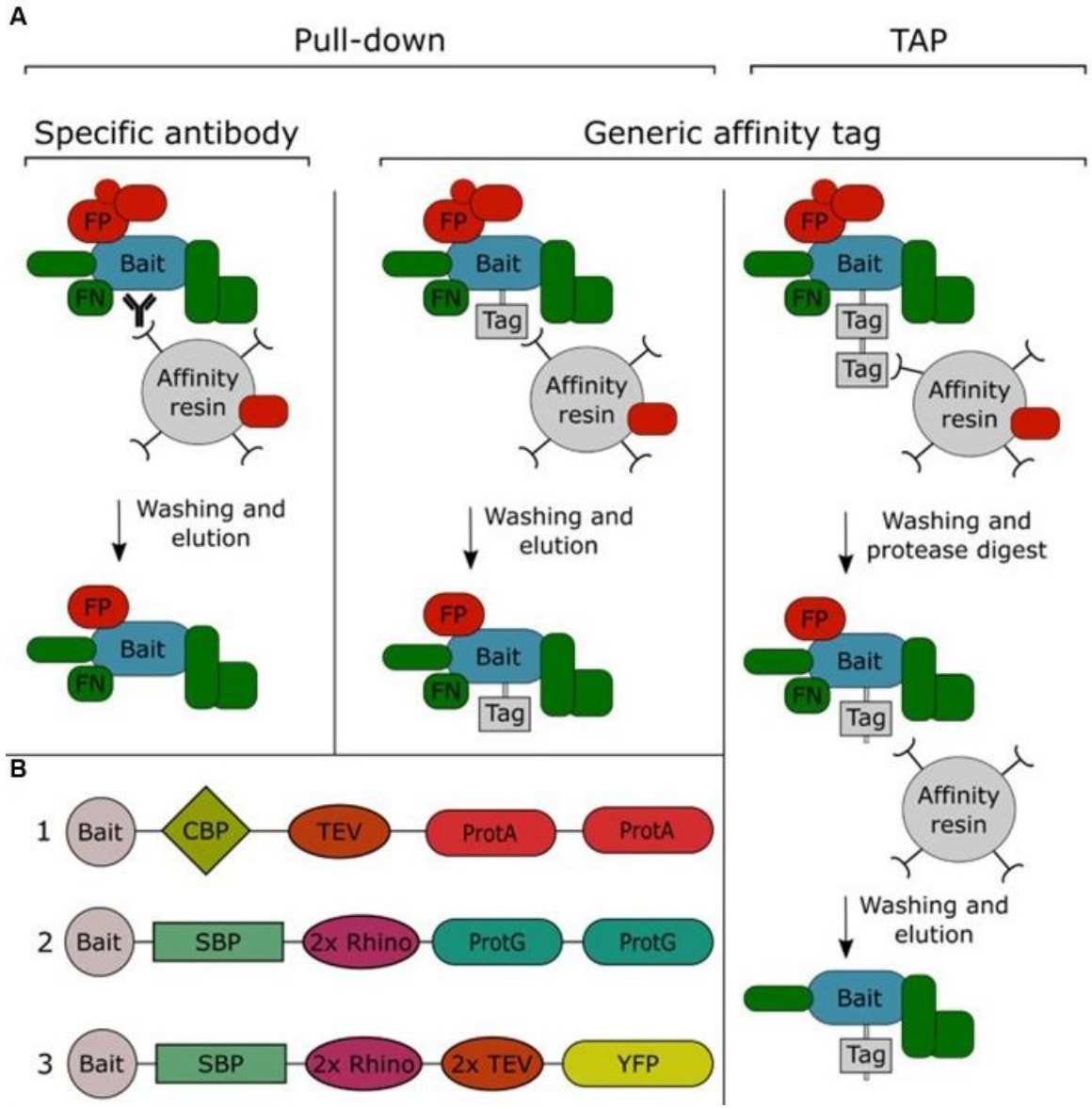


Figure 2: Overview of the different AP/MS approaches (A) and available TAP tags (B). TAP tags: (1) TAPi tag; (2) GS^{rhino} tag; (3) GS^{yellow} tag. CBP, calmodulin binding protein; FP, false positive; FN, false negative; ProtA, protein A domain; ProtG, protein G domain; 2x Rhino, double recognition site for the Rhinovirus 3C protease; TEV, recognition site for the tobacco etch virus protease; SBP, streptavidin-binding peptide; YFP, yellow fluorescent protein. Figure adopted from Bontinck et al. (2018).

Biotin-based proximity labeling

An alternative approach that is able to identify stable and transient PPIs, even for low abundant proteins, is the proximity-dependent labeling method (Roux, 2013; Rees et al., 2015). This technique relies on an enzyme, typically a biotin ligase, that is capable of covalently labeling, e.g. biotinylation, a protein in the immediate vicinity. By fusing this enzyme to a bait protein, interacting partners will be covalently labeled *in vivo*, independent of their binding affinity. Subsequent pull-down and MS analysis will lead to the identification of the labeled proteins. Enzymes that are typically used for proximity labeling are peroxidase enzymes, like ascorbate peroxidase (APEX) (Rhee et al., 2013), and biotin ligases, like BirA (Roux et al., 2012). Both label neighboring proteins with biotin, allowing a subsequent stringent pull-down with streptavidin. Engineering these proximity labeling enzymes have led to the development of more reliable, more efficient and faster biotinylating enzymes, like TurboID (Branon et al., 2018). This mutant BirA enzyme reduces the labeling time from more than 18 hours to 10 minutes using a lower amount of biotin. This improvement in time resolution is important for analyzing PPI dynamics and identifying proteins that turn over rapidly *in vivo*.

Application of proximity labeling in plants remains limited with studies in rice protoplasts, tobacco and *Arabidopsis* (Lin et al., 2017; Conlan et al., 2018; Khan et al., 2018; Das et al., 2019). However, with the development of TurboID, applications in plants are facilitated. Recently the first *in planta* TurboID study has been published, revealing the identification of new putative co-activator and -repressor complex components for a transcription factor in young guard cells (Mair et al., 2019). They showed that TurboID-based proximity labeling can be applied under normal plant growth conditions and that labeling times of under 10 min can give immunoblot-detectable signals. However, they also suggest that longer incubation may be required for protein identification by MS. Indeed, TurboID has been applied in *Arabidopsis* for mapping the signaling network of the GSK3 kinase, by incubating transgenic seedling for at least one hour with low concentration of biotin (Tae-Wuk Kim, 2019). In addition, TurboID has also been applied in *Nicotiana benthamiana* leafs, *Arabidopsis thaliana* cell suspension culture and *Solanum spp.* hairy roots for capturing membrane protein interactomes and the octameric endocytic TPLATE complex (Deepanksha Arora, 2019).

Results obtained with biotin-based proximity labeling in plants need to be carefully interpreted. Plants synthesize biotin, histones are biotinylated endogenously, and biotin is used as an essential co-factor for carboxylases (Nikolau et al., 1985; Nikolau et al., 2003; Zempleni et al., 2009). Endogenous biotin and biotinylated proteins will

contaminate the streptavidin-based pull-down and will lead to a low signal-to-noise ratio, making it more challenging to identify bona-fide interaction partners with MS. In addition, account must be taken of the fact that proximity labeling will also result in the labeling of non-interacting proteins that are in the neighborhood of the bait protein. Therefore, biotin-based proximity labeling in plants should always be combined with a quantitative MS analysis, comparing the results relative to a negative control, being a wild type plant (cell), unrelated bait protein with similar localization and/or an untreated (no addition of biotin) sample (Mair et al., 2019).

Identification of transient protein-protein interactions

As stated before, protein complexes often have a set of more stably interacting proteins as well as more unstable or transient interactions. Studying these transient interactions by applying affinity purification is challenging as most of the time only the stable components are retrieved and the weak, transient interactions are lost. Therefore, binary PPI tools are to date, the most appropriate techniques to identify transient interactions. Y2H has the advantage that a transcriptional readout leads to an amplification of the response upon interaction, which offers more sensitivity for transient interactions. To enrich for transient interactions during a Y2H analysis, repeated screens are necessary as these interactions do not occur in every single screen (Vinayagam et al., 2010). Furthermore, BiFC studies result in a strong stabilization of the fluorescent protein and therefore also stabilizes weak transient PPIs. However, necessary adjustments to the binary PPI systems often do not take transient PPIs into account. For example, combining two reporter genes to increase the stringency of the Y2H system, leads to a decrease in detection of false positives, but also of weak, transient interactions. In addition, PTMs are often essential for transient interactions and are frequently missed when analyzed in yeast.

To be able to identify weak, transient PPIs related to multiprotein complexes *in vivo*, adjustments have been made to the affinity-based methods implementing chemical cross-linking to fixate PPIs (Rohila et al., 2004; Tagwerker et al., 2006; Stingl et al., 2008; Van Leene et al., 2019). The resulting covalent bonds between interacting proteins ensures that transient interactions are not lost during the necessary washing steps. For example, this strategy has been applied in plants to unravel the interactome of membrane proteins (Pertl-Obermeyer et al., 2014; Bellati et al., 2016). Cross-linking-based AP in combination with quantitative MS can help to define the *in vivo* stable and transient protein interactions. This has been demonstrated in yeast where a technology called transient isotopic differentiation of interactions as random or targeted (transient I-DIRT) was developed (Smart et al., 2009; Byrum et al., 2012). One culture of

isotopically light cells containing the affinity-tagged bait protein and a second negative control culture of isotopically heavy cells, are subjected to *in vivo* cross-linking and are mixed in a ratio of 1:1. After affinity purification and trypsin digestion, stable interacting proteins will be identified by MS as isotopically light peptides, while contaminants will have a 1:1 mixture of light and heavy peptides. Transient interacting proteins will show an intermediate level of isotopically light peptides.

Although *in vivo* cross-linking is beneficial for the identification of transient interactions, it also results in a higher number of false positives. In addition, it only reveals PPIs at the time point of cross-linking, limiting the amount of co-purified interacting partners. In that respect, proximity labeling seems to be more ideal to capture all interacting proteins, including stable, transient and dynamic interactions. Further on, interaction studies are not sufficient for the identification of substrates for a specific enzyme or protein complex. To make sure that the identified PPI also results in bait-specific prey alterations, further validation of newly identified preys is needed. Although, implementation of stimuli or modified PPI analysis conditions, can already hint towards bait specific substrates. For example, biotin-based proximity labeling was used to reveal stable interacting proteins and transient interacting substrates for the $SCF^{\beta\text{-}TrCP1/2}$ E3 ligase by combining proximity labeling and MG132 treatment (Coyaud et al., 2015). Because E3 ligase substrates are rapidly degraded by the 26S proteasome upon ubiquitination, these proteins are generally missed during AP-MS studies. Combining 26S proteasome inhibitor (MG132) treatment with biotin-based proximity labeling and semi quantitative mass spectrometry, enabled Coyaud et al. for enrichment of $SCF^{\beta\text{-}TrCP1/2}$ E3 substrates in human cells.

Alternatively, a substrate trapping approach has been applied for the identification of RING-type E3 ligases substrates (Pauwels et al., 2015; Nagels Durand et al., 2016). For this strategy, mutations have been introduced in the E3 ligase disrupting the RING domain. As a consequence, the interaction with the Ub-charged E2-conjugating enzyme is lost, while the interaction with substrates remain. However these substrates are not ubiquitinated and are not targeted for degradation, resulting in the enrichment of the E3 target proteins during subsequent affinity purification. This approach has led to the identification of GRXS17 as an E3 substrate of the heterodimeric RGLG3 and RGLG4 complex in *Arabidopsis* (Nagels Durand et al., 2016).

Identification of kinase substrates with AP-MS

Recently we published the phosphorylation and interaction landscape of the plant target of rapamycin (TOR) kinase, including some known and newly identified direct substrates (Van Leene et al., 2019). TOR is an evolutionarily conserved serine/threonine protein kinase, which coordinate cell growth based on energy and nutrient availability. The complex consists of three core subunits, TOR, RAPTOR and LST8. The list of known plant TOR substrates is limited, including the S6 kinase (S6K), which stimulates protein synthesis through phosphorylation of ribosomal S6 proteins, and the protein phosphatase 2A regulatory subunit TAP46, which regulates cell growth in coordination with nutrient and environmental conditions (Dobrenel et al., 2016). To further unravel the upstream and downstream network components of the plant TOR pathway, we performed a systematic phosphoproteomics screen and affinity purification assays in *Arabidopsis* cell suspension cultures (PSB-D). PSB-D can be synchronized through sucrose starvation and repletion (Menges and Murray, 2002), which we used to enrich for sugar-dependent TOR signaling events. Sucrose was added to 24 hours sucrose-starved cell cultures, whether or not pretreated 2 hours before sucrose repletion with the highly specific ATP-competitive TOR inhibitor AZD8055 or with the allosteric TOR inhibitor rapamycin. The corresponding protein extracts were digested with trypsin/Lys-C, and phosphopeptides were enriched using TiO2 magnetic beads that were modified with lactic acid to reduce non-specific binding of acidic peptides. Enriched phosphopeptides were analyzed by label-free quantitative MS, giving rise to a filtered data set of 4,988 phosphopeptides on 2,119 proteins. To select for TOR dependent phosphopeptides, we looked at phosphorylation dynamics at different time points, with or without TOR inhibitor treatments. In total, 111 unambiguous TOR-dependent phosphosites were detected on 83 proteins. To further elucidate which TOR-regulated phosphoproteins were direct substrates of the TOR complex, TOR kinase protein interactions were characterized by AP-MS, using TOR, LST8-1 and Raptor1B as bait proteins. At first we applied TAP, using the GS^{rhino}-tag and GS^{yellow}-tag, revealing interactions with 53 proteins. When we compared this TOR interactome with the TOR-dependent phosphoproteome, only one protein was found in the overlap, being eIF2B-δ1, a translation initiation factor. This low overlap indicates that the TAP analysis mainly identified stable protein interactions, whereas kinase–substrate interactions, which are more transient in nature, were probably lost during the long double-step TAP purification. Therefore, we tested a one-step pull-down protocol, derived from our GS^{rhino} TAP method (figure 3). The bait proteins, LST8-1 and Raptor1B fused to GS^{rhino}, were trapped on home-made magnetic IgG beads which contain high binding capacity. These magnetic beads allowed us to purify protein complexes using the strong affinity between ProtG and IgG. It also resulted in faster,

integral washing steps, and shorter incubation times. Due to these advantages, we could transform our relative slow TAP technique to a faster pull-down technique (AP-MS). To reveal specific interactors of LST8-1 and RAPTOR1B, a label-free quantitative MS was performed by comparison to an equivalent control data set. AP-MS revealed 119 interacting partners of the TOR complex, of which 8 were also TOR-dependent phosphorylated. From these 8 putative TOR substrates, three (AML5, ATG1b and ATG13) are orthologues of known TOR substrates in yeast or mammalian cells. Several putative plant-specific TOR substrates were identified as well, including a protein that is upregulated during senescence (S40-7) and the translation initiation factor eIF2B-δ1. To further validate these putative new TOR substrates, we performed *in vitro* kinase assays revealing the TOR dependent phosphorylation of ATG13, S40-7 and eIF2B-δ1. These results show that our AP-MS strategy was more efficient than TAP for the discovery of direct TOR substrates.

The PSB-D sucrose-dependent phoshoproteome that we generated also revealed a large amount of proteins which phosphorylation status diminished upon sucrose repletion. Further analysis uncovered that a relative high percentage of the identified phosphorylation sites resemble the consensus sequence for phosphorylation by the SnRK1 kinase complex (unpublished data). SnRK1, just as TOR, is an evolutionarily conserved protein kinase essential for plant growth. However, while TOR is activated in nutrient-rich conditions to promote growth, SnRK1 is activated by energy deficit and will lead to growth arrest as long as growth conditions are not beneficial. So it is no coincidence that we find known (e.g. bZIP transcription factors) and putative SnRK1 substrates in the phosphoproteome of 24 hours sucrose-starved cell cultures. To further elucidate the direct SnRK1 substrates, we performed AP-MS on two of the three subunits, being KIN10 and SnRK1βγ. This resulted in the identification of known SnRK1 regulators and targets, including TOR, as well as novel interacting partners (unpublished data). Ongoing experiments, among which *in vitro* kinase assays and transient expression assays, will further validate the AP-MS results and will reveal if direct SnRK1 substrates have been purified with our AP-MS tool.

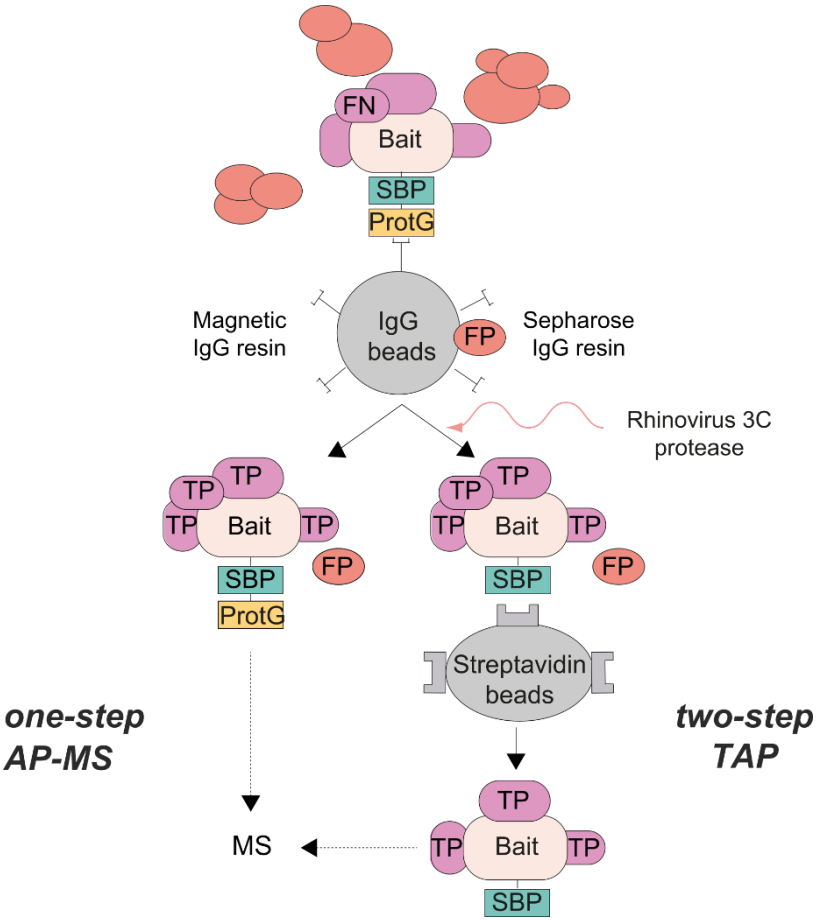


Figure 3: Schematic overview of the GS^{rhino}-based pull-down and TAP protocols. Protein complexes incorporating the GS^{rhino}-tagged bait protein are isolated using either TAP or a one-step pull-down strategy. In the pull-down experiment, protein complexes are captured during a 45 min incubation with home-made IgG magnetic beads. During TAP, protein complexes are first captured using IgG sepharose beads (1 h), specifically eluted by Rhinovirus 3C protease cleavage (1 h) and further purified by incubation with Streptavidin sepharose beads (1 h). SBP = streptavidin-binding peptide, ProtG = ProteinG tag, FN = false negative interaction here depicted as an interaction specifically lost during TAP, TP = true positive interactions, FP = false positive interactions, MS = mass spectrometry. Figure adopted from Van Leene et al. (2019).

Summary and perspectives

Proteins are the workhorses of a biological system and mutual protein interactions are essential in a living organism. However, the different functionalities, dynamics and complexity of PPIs, makes it a challenging field to investigate. The list of PPI analysis tools is very extensive, and only a small portion, relevant to plant research, has been discused here. We made the distinction between *in situ* binary methods and affinity-based methods, which have been succesfully applied in plants.

The most widley applied binary system is Y2H, and has been modified numerous times to circumvent some of the limitations associated with the standard Y2H strategy. Alternative two-hybrid systems have been reviewed in detail by Stynen and co-workers (Stynen et al., 2012), including membrane associated Y2H systems and implementation of different transcription activators or repressors. However, in relationship to plant PPIs, Y2H remains suboptimal. Plant interactions are analyzed outside the native cellular context, leading to a low signal-to-noise ratio. Plant PPIs often are dependent on PTMs, like tyrosine phosphorylation, a mechanism absent in yeast. To obtain more relevant data, the Y2H system has been adopted in *Arabidopsis*

protoplasts (P2H) (Ehlert et al., 2006; Li et al., 2014). However, to be able to study PPI dynamics and/or their localization, versatile techniques have been developed using bioluminescence (split-luciferase, BRET) or fluorescence (BiFC, FRET). These methods allow for a straightforward visualization of binary interactions in living cells and in real time, independent of the PPI localization or affinity strength. These techniques have become increasingly popular in the plant field with applications in diverse targets, ranging from protoplasts to seedlings, leaves, or epidermal cells in *Arabidopsis* but also in tobacco, mustard, parsley, leek, and onion plants (Ohad et al., 2007; Ohad and Yalovsky, 2010). However, it must be taken in to account that binary PPI studies are prone for false positives and false negatives, making it necessary to validate binary PPIs with complementary assays. Validation can be achieved by performing several different PPI techniques, however, it has been shown that performing one assay under different experimental conditions, e.g. the same binary PPI assay in different organisms or with different fusion orientations on bait and prey protein, can reveal true PPI in an effortless manner (Choi et al., 2019).

Athough high-throughput Y2H screens have been performed to map plant binary protein interactomes (Erffelinck et al., 2018), more relevant data on a genome-wide level is obtained with affinity-based methods. Preferably, immunoprecipitation is applied on an endogenous expressed protein of interest with a specific antibody, followed by quantitative mass spectrometry to identify the interacting partners. Unfortunetaly, the amount of plant specific antibodies is limited, causing the need for the transgenic expression of a tagged bait protein. Preference is given to the overexpression of the tagged bait protein, which favours competition with the endogenous counterpart for PPIs. Expression under the endogenous promoter is also a possibility, but competition with the endogenous counterpart should be avoided by using a knock out mutant (Dedecker et al., 2015). A third option could be the insertion of the tag in the endogenous bait locus, allowing near-to-physiological expression. While this strategy can be applied in yeast via homologous recombination, application in plants remains to date inefficient, although attempts have been made with the CRISPR system (Puchta, 2017). Furthermore, the affinity tag can obstruct the functionality of the bait protein, cause steric hindrance and interfere with certain protein interactions. Therfore, it is recommended to test N- and C-terminal fusions in parallel, which also applies to the binary PPI methods (Dedecker et al., 2015; Choi et al., 2019).

Affinity purification mainly results in the identification of stable PPIs because these interactions do not disassemble during the different protein extraction and purification steps. To include the purification of weak and transient interactions, *in vivo* cross-linking can be implemented, but this is accompanied by a higher number of false

positives. Another option to pull-down weak, transient interactions is by shortening the affinity purification protocol, as we have done for the identification of kinase substrates (Van Leene et al., 2019). Our AP-MS protocol combines the strong affinity of ProtG with IgG and the high binding capacity of magnetic IgG beads to pull-down a bait protein and its interacting partners in approximately three hours. In addition, the ongoing optimization of MS technologies, advanced mathematical modeling, and data-driven elimination of common contaminants, increases the reliability of AP-MS results. We demonstrated that our AP-MS strategy was more efficient than TAP for the discovery of direct TOR substrates and promising results have been obtained for the SnRK1 interactome (unpublished data). However, the overlap with the TOR dependent phosphoproteome remained low, and known TOR substrates, like S6K and TAP46, were not retrieved by AP-MS, nor with TAP. We attempted to increase the overlap by performing pull-downs on sucrose synchronized cell cultures, maximizing the TOR activity and its interactions with substrates, but this did not give a better result (unpublished data). Because of the practical limitations and the necessary washing steps, AP-MS will remain a suboptimal technique for identification of weak and transient interactions.

A promising new PPI tool that avoids the need for retaining interactions during the different affinity purification steps, is proximity labeling. Especially with the development of TurboID, a more reliable, more efficient and faster biotin ligase (Branon et al., 2018), proximity labeling will provide the next generation of PPI tools in plants allowing *in vivo* biotin labeling of all proteins in close proximity of the bait protein, including weak and transient interactions. While affinity-based methods pull-down PPIs at a certain time point, proximity labeling has the advantage of accumulating labeled proteins over the time period of biotin treatment, leading to the capture of dynamic interactions. In addition, proximity labeling also occurs with endogenous biotin over the growth period of the plant, creating a large pool of biotinylated protein interactors which are purified with high affinity for streptavidin and identified by MS. However, the same biotinylation period allows non-interacting cellular proteins to enter the labeling radius of TurboID, resulting in the accumulation of false positives. Therefore, similar to AP-MS, a quantitative MS analysis relative to a proper negative control is necessary (Mair et al., 2019). Furthermore, the relative large TurboID enzyme (35 kDa) can hinder PPIs, and when analyzing multiprotein complexes, some interacting partners, hidden in the core of the complex, could be inaccessible for biotin labeling, leading to false negative results.

The list of PPI tools is extensive, but no single PPI assay is superior to any other, including the most recently developed technologies. All methods have their

advantages and disadvantages. Depending on the aim and the characteristics of the bait protein, one method will be more suitable than the other. A well-considered choice must be made, combining different complementary assays followed by *in vivo* validation.

References

Alonso-Lopez D, Campos-Laborie FJ, Gutierrez MA, Lambourne L, Calderwood MA, Vidal M, De Las Rivas J (2019) APID database: redefining protein-protein interaction experimental evidences and binary interactomes. Database (Oxford) **2019**

Backstrom S, Elfving N, Nilsson R, Wingsle G, Bjorklund S (2007) Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. Mol Cell **26:** 717-729

Bellati J, Champeyroux C, Hem S, Rofidal V, Krouk G, Maurel C, Santoni V (2016) Novel Aquaporin Regulatory Mechanisms Revealed by Interactomics. Mol Cell Proteomics **15:** 3473-3487

Berendzen KW, Bohmer M, Wallmeroth N, Peter S, Vesic M, Zhou Y, Tiesler FK, Schleifenbaum F, Harter K (2012) Screening for in planta protein-protein interactions combining bimolecular fluorescence complementation with flow cytometry. Plant Methods **8:** 25

Berry LD, Gould KL (1996) Regulation of Cdc2 activity by phosphorylation at T14/Y15. Prog Cell Cycle Res **2:** 99-105

Besbrugge N, Van Leene J, Eeckhout D, Cannoot B, Kulkarni SR, De Winne N, Persiau G, Van De Slijke E, Bontinck M, Aesaert S, Impens F, Gevaert K, Van Damme D, Van Lijsebettens M, Inze D, Vandepoele K, Nelissen H, De Jaeger G (2018) GS(yellow), a Multifaceted Tag for Functional Protein Analysis in Monocot and Dicot Plants. Plant Physiol **177:** 447-464

Bhat RA, Lahaye T, Panstruga R (2006) The visible touch: in planta visualization of protein-protein interactions by fluorophore-based methods. Plant Methods **2:** 12

Bontinck M, Van Leene J, Gadeyne A, De Rybel B, Eeckhout D, Nelissen H, De Jaeger G (2018) Recent Trends in Plant Protein Complex Analysis in a Developmental Context. Front Plant Sci **9:** 640

Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inze D, De Veylder L, Russinova E (2010) Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network. Plant Cell **22:** 1264-1280

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Braun P, Carvunis AR, Charloteaux B, Dreze M, Ecker JR, Hill DE, Roth FP, Vidal M, Galli M, Balumuri P, Bautista V, Chesnut JD, Kim RC, de los Reyes C, Gilles P, Kim CJ, Matrubutham U, Mirchandani J, Olivares E, Patnaik S, Quan R, Ramaswamy G, Shinn P, Swamilingiah GM, Wu S, Ecker JR, Dreze M, Byrdsong D, Dricot A, Duarte M, Gebreab F, Gutierrez BJ, MacWilliams A, Monachello D, Mukhtar MS, Poulin MM, Reichert P, Romero V, Tam S, Waaijers S, Weiner EM, Vidal M, Hill DE, Braun P, Galli M, Carvunis AR, Cusick ME, Dreze M, Romero V, Roth FP, Tasan M, Yazaki J, Braun P, Ecker JR, Carvunis AR, Ahn YY, Barabasi AL, Charloteaux B, Chen HM, Cusick ME, Dangl JL, Dreze M, Ecker JR, Fan CY, Gai LT, Galli M, Ghoshal G, Hao T, Hill DE, Lurin C, Milenkovic T, Moore J, Mukhtar MS, Pevzner SJ, Przulj N, Rabello S, Rietman EA, Rolland T, Roth FP, Santhanam B, Schmitz RJ, Spooner W, Stein J, Tasan M, Vandenhaute J, Ware D, Braun P, Vidal M, Braun P, Carvunis AR, Charloteaux B, Dreze M, Galli M, Vidal M, Co AIM (2011) Evidence for Network Evolution in an Arabidopsis Interactome Map. Science **333:** 601-607

Bucherl CA, Bader A, Westphal AH, Laptenok SP, Borst JW (2014) FRET-FLIM applications in plant systems. Protoplasma **251:** 383-394

Burkle L, Meyer S, Dortay H, Lehrach H, Heyl A (2005) In vitro recombination cloning of entire cDNA libraries in Arabidopsis thaliana and its application to the yeast two-hybrid system. Funct Integr Genomics **5:** 175-183

Byrum S, Smart SK, Larson S, Tackett AJ (2012) Analysis of stable and transient protein-protein interactions. Methods Mol Biol **833:** 143-152

Causier B (2004) Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom Rev **23:** 350-367

Causier B, Davies B (2002) Analysing protein-protein interactions with the yeast two-hybrid system. Plant Mol Biol **50:** 855-870

Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res **41:** D816-823

Choi SG, Olivet J, Cassonnet P, Vidalain PO, Luck K, Lambourne L, Spirohn K, Lemmens I, Dos Santos M, Demeret C, Jones L, Rangarajan S, Bian W, Coutant EP, Janin YL, van der Werf S, Trepte P, Wanker EE, De Las Rivas J, Tavernier J, Twizere JC, Hao T, Hill DE, Vidal M, Calderwood MA, Jacob Y (2019) Maximizing binary interactome mapping with a minimal number of assays. Nat Commun **10:** 3907

Conlan B, Stoll T, Gorman JJ, Saur I, Rathjen JP (2018) Development of a Rapid in planta BioID System as a Probe for Plasma Membrane-Associated Immunity Proteins. Front Plant Sci **9:** 1882

Coriano C, Powell E, Xu W (2016) Monitoring Ligand-Activated Protein-Protein Interactions Using Bioluminescent Resonance Energy Transfer (BRET) Assay. Methods Mol Biol **1473:** 3-15

Coyaud E, Mis M, Laurent EM, Dunham WH, Couzens AL, Robitaille M, Gingras AC, Angers S, Raught B (2015) BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates. Mol Cell Proteomics **14:** 1781-1795

Dale R, Ohmuro-Matsuyama Y, Ueda H, Kato N (2016) Mathematical Model of the Firefly Luciferase Complementation Assay Reveals a Non-Linear Relationship between the Detected Luminescence and the Affinity of the Protein Pair Being Analyzed. PLoS One **11:** e0148256

Das PP, Macharia MW, Lin Q, Wong SM (2019) In planta proximity-dependent biotin identification (BioID) identifies a TMV replication co-chaperone NbSGT1 in the vicinity of 126kDa replicase. J Proteomics **204:** 103402

Day RN, Davidson MW (2009) The fluorescent protein palette: tools for cellular imaging. Chem Soc Rev **38:** 2887-2921

De Bondt HL, Rosenblatt J, Jancarik J, Jones HD, Morgan DO, Kim SH (1993) Crystal structure of cyclin-dependent kinase 2. Nature **363:** 595-602

De Rybel B, Moller B, Yoshida S, Grabowicz I, Barbier de Reuille P, Boeren S, Smith RS, Borst JW, Weijers D (2013) A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. Dev Cell **24:** 426-437

Debernardi JM, Mecchia MA, Vercruyssen L, Smaczniak C, Kaufmann K, Inze D, Rodriguez RE, Palatnik JF (2014) Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. Plant J **79:** 413-426

Dedecker M, Van Leene J, De Jaeger G (2015) Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry. Curr Opin Plant Biol **24:** 1-9

Dedecker M, Van Leene J, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Cannoot B, Vercruysse L, Dumoulin L, Wojsznis N, Gevaert K, Vandenabeele S, De Jaeger G (2016) Transferring an optimized TAP-toolbox for the isolation of protein complexes to a portfolio of rice tissues. Plant Mol Biol **91:** 341-354

Deepanksha Arora NBA, Chen Liu, Petra Van Damme, Lam Dai Vu, Anna Tornkvist, Francis Impens, Dominique Eeckhout, Alain Goossens, Geert De Jaeger, Thomas Ott, Panagiotis Moschou, Daniel Van Damme (2019) Establishment of Proximity-dependent Biotinylation Approaches in Different Plant Model Systems. BioRxiv

Dobrenel T, Caldana C, Hanson J, Robaglia C, Vincentz M, Veit B, Meyer C (2016) TOR Signaling and Nutrient Sensing. Annu Rev Plant Biol **67:** 261-285

Dohmen RJ, Madura K, Bartel B, Varshavsky A (1991) The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. Proc Natl Acad Sci U S A **88:** 7351-7355

Draetta GF (1997) Cell cycle: will the real Cdk-activating kinase please stand up. Curr Biol **7:** R50-52

Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Droge-Laser W (2006) Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. Plant J **46:** 890-900

Elledge SJ, Harper JW (1994) Cdk inhibitors: on the threshold of checkpoints and development. Curr Opin Cell Biol **6:** 847-852

Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GT, Inze D, Ferreira PC (2011) The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. Plant J **68:** 351-363

Eloy NB, Gonzalez N, Van Leene J, Maleux K, Vanhaeren H, De Milde L, Dhondt S, Vercruysse L, Witters E, Mercier R, Cromer L, Beemster GTS, Remaut H, Van Montagu MCE, De Jaeger G, Ferreira PCG, Inze D (2012) SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization. Proceedings of the National Academy of Sciences of the United States of America **109:** 13853-13858

Eloy NB, Lima MD, Ferreira PCG, Inze D (2015) The Role of the Anaphase-Promoting Complex/Cyclosome in Plant Growth. Critical Reviews in Plant Sciences **34:** 487-505

Erffelinck ML, Ribeiro B, Perassolo M, Pauwels L, Pollier J, Storme V, Goossens A (2018) A user-friendly platform for yeast two-hybrid library screening using next generation sequencing. PLoS One **13:** e0201270

Fiedorczuk K, Letts JA, Degliesposti G, Kaszuba K, Skehel M, Sazanov LA (2016) Atomic structure of the entire mammalian mitochondrial complex I. Nature **538:** 406-410

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature **340:** 245-246

Fujikawa Y, Kato N (2007) Split luciferase complementation assay to study protein-protein interactions in Arabidopsis protoplasts. Plant J **52:** 185-195

Fujikawa Y, Nakanishi T, Kawakami H, Yamasaki K, Sato MH, Tsuji H, Matsuoka M, Kato N (2014) Split luciferase complementation assay to detect regulated protein-protein interactions in rice protoplasts in a large-scale format. Rice (N Y) **7:** 11

Gadeyne A, Sanchez-Rodriguez C, Vanneste S, Di Rubbo S, Zauber H, Vanneste K, Van Leene J, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Cannoot B, Vercruysse L, Mayers JR, Adamowski M, Kania U, Ehrlich M, Schweighofer A, Ketelaar T, Maere S, Bednarek SY, Friml J, Gevaert K, Witters E, Russinova E, Persson S, De Jaeger G, Van Damme D (2014) The TPLATE adaptor complex drives clathrin-mediated endocytosis in plants. Cell **156:** 691-704

Geley S, Kramer E, Gieffers C, Gannon J, Peters JM, Hunt T (2001) Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. J Cell Biol **153:** 137-148

Gerace E, Moazed D (2015) Affinity Purification of Protein Complexes Using TAP Tags. Methods Enzymol **559:** 37-52

Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. Nature **349:** 132-138

Goossens J, De Geyter N, Walton A, Eeckhout D, Mertens J, Pollier J, Fiallos-Jurado J, De Keyser A, De Clercq R, Van Leene J, Gevaert K, De Jaeger G, Goormachtig S, Goossens A (2016) Isolation of protein complexes from the model legume Medicago truncatula by tandem affinity purification in hairy root cultures. Plant J **88:** 476-489

Gyuris J, Golemis E, Chertkov H, Brent R (1993) Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell **75:** 791-803

Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giege P (2011) An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. Plant Cell **23:** 730-740

Hirst M, Ho C, Sabourin L, Rudnicki M, Penn L, Sadowski I (2001) A two-hybrid system for transactivator bait proteins. Proc Natl Acad Sci U S A **98:** 8726-8731

Horstman A, Tonaco IA, Boutilier K, Immink RG (2014) A cautionary note on the use of split-YFP/BiFC in plant protein-protein interaction studies. Int J Mol Sci **15:** 9628-9643

Hu CD, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. Nat Biotechnol **21:** 539-545

Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature **376:** 313-320

Johnsson N, Varshavsky A (1994) Split ubiquitin as a sensor of protein interactions in vivo. Proc Natl Acad Sci U S A **91:** 10340-10344

Jones AM, Xuan Y, Xu M, Wang RS, Ho CH, Lalonde S, You CH, Sardi MI, Parsa SA, Smith-Valle E, Su T, Frazer KA, Pilot G, Pratelli R, Grossmann G, Acharya BR, Hu HC, Engineer C, Villiers F, Ju C, Takeda K, Su Z, Dong Q, Assmann SM, Chen J, Kwak JM, Schroeder JI, Albert R, Rhee SY, Frommer WB (2014) Border control--a membrane-linked interactome of Arabidopsis. Science **344:** 711-716

Karcher RL, Deacon SW, Gelfand VI (2002) Motor-cargo interactions: the key to transport specificity. Trends Cell Biol **12:** 21-27

Kaspar M, Dienemann A, Schulze C, Sprenger F (2001) Mitotic degradation of cyclin A is mediated by multiple and novel destruction signals. Curr Biol **11:** 685-690

Kato N, Fujikawa Y, Fuselier T, Adamou-Dodo R, Nishitani A, Sato MH (2010) Luminescence detection of SNARE-SNARE interaction in Arabidopsis protoplasts. Plant Mol Biol **72:** 433-444

Khan M, Youn JY, Gingras AC, Subramaniam R, Desveaux D (2018) In planta proximity dependent biotin identification (BioID). Sci Rep **8:** 9212

Kimple ME, Brill AL, Pasker RL (2013) Overview of affinity tags for protein purification. Curr Protoc Protein Sci **73:** Unit 9 9

Kodama Y, Wada M (2009) Simultaneous visualization of two protein complexes in a single plant cell using multicolor fluorescence complementation analysis. Plant Mol Biol **70:** 211-217

Kramer ER, Scheuringer N, Podtelejnikov AV, Mann M, Peters JM (2000) Mitotic regulation of the APC activator proteins CDC20 and CDH1. Mol Biol Cell **11:** 1555-1569

Lalonde S, Sero A, Pratelli R, Pilot G, Chen J, Sardi MI, Parsa SA, Kim DY, Acharya BR, Stein EV, Hu HC, Villiers F, Takeda K, Yang Y, Han YS, Schwacke R, Chiang W, Kato N, Loque D, Assmann SM, Kwak JM, Schroeder JI, Rhee SY, Frommer WB (2010) A membrane protein/signaling protein interaction network for Arabidopsis version AMPv2. Front Physiol **1:** 24

Laser H, Bongards C, Schuller J, Heck S, Johnsson N, Lehming N (2000) A new screen for protein interactions reveals that the Saccharomyces cerevisiae high mobility group proteins Nhp6A/B are involved in the regulation of the GAL1 promoter. Proc Natl Acad Sci U S A **97:** 13732-13737

Li CH, Chiang CP, Yang JY, Ma CJ, Chen YC, Yen HE (2014) RING-type ubiquitin ligase McCPN1 catalyzes UBC8-dependent protein ubiquitination and interacts with Argonaute 4 in halophyte ice plant. Plant Physiol Biochem **80:** 211-219

Li JF, Bush J, Xiong Y, Li L, McCormack M (2011) Large-scale protein-protein interaction analysis in Arabidopsis mesophyll protoplasts by split firefly luciferase complementation. PLoS One **6:** e27364

Li Y (2011) The tandem affinity purification technology: an overview. Biotechnol Lett **33:** 1487-1499

Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif **41:** 98-105

Lin Q, Zhou Z, Luo W, Fang M, Li M, Li H (2017) Screening of Proximal and Interacting Proteins in Rice Protoplasts by Proximity-Dependent Biotinylation. Front Plant Sci **8:** 749

Lin T, Scott BL, Hoppe AD, Chakravarty S (2018) FRETting about the affinity of bimolecular protein-protein interactions. Protein Sci **27:** 1850-1856

Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L (2005) Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. J Am Chem Soc **127:** 146-157

Mair A, Xu SL, Branon TC, Ting AY, Bergmann DC (2019) Proximity labeling of protein complexes and cell type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife **8**

Marsolier MC, Prioleau MN, Sentenac A (1997) A RNA polymerase III-based two-hybrid system to study RNA polymerase II transcriptional regulators. J Mol Biol **268:** 243-249

Masters SC (2004) Co-immunoprecipitation from transfected cells. Methods Mol Biol **261:** 337-350

Matiolli CC, Melotto M (2018) A Comprehensive Arabidopsis Yeast Two-Hybrid Library for Protein-Protein Interaction Studies: A Resource to the Plant Research Community. Mol Plant Microbe Interact **31:** 899-902

Menges M, Murray JA (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J **30:** 203-212

Meyer K, Selbach M (2015) Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. Front Genet **6:** 237

Miller KE, Kim Y, Huh WK, Park HO (2015) Bimolecular Fluorescence Complementation (BiFC) Analysis: Advances and Recent Applications for Genome-Wide Interaction Studies. J Mol Biol **427:** 2039-2055

Mitsuda N, Ikeda M, Takada S, Takiguchi Y, Kondou Y, Yoshizumi T, Fujita M, Shinozaki K, Matsui M, Ohme-Takagi M (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in Arabidopsis thaliana. Plant Cell Physiol **51:** 2145-2151

Mockli N, Deplazes A, Hassa PO, Zhang Z, Peter M, Hottiger MO, Stagljar I, Auerbach D (2007) Yeast split-ubiquitin-based cytosolic screening system to detect interactions between transcriptionally active proteins. Biotechniques **42:** 725-730

Mravec J, Petrasek J, Li N, Boeren S, Karlova R, Kitakura S, Parezova M, Naramoto S, Nodzynski T, Dhonukshe P, Bednarek SY, Zazimalova E, de Vries S, Friml J (2011) Cell plate restricted association of DRP1A and PIN proteins is required for cell polarity establishment in Arabidopsis. Curr Biol **21:** 1055-1060

Nagels Durand A, Inigo S, Ritter A, Iniesto E, De Clercq R, Staes A, Van Leene J, Rubio V, Gevaert K, De Jaeger G, Pauwels L, Goossens A (2016) The Arabidopsis Iron-Sulfur Protein GRXS17 is a Target of the Ubiquitin E3 Ligases RGLG3 and RGLG4. Plant Cell Physiol **57:** 1801-1813

Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, Van Lijsebettens M, Goormachtig S, Vandepoele K, Van Leene J, Muszynski M, Gevaert K, Inze D, De Jaeger G (2015) Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf. Plant Cell **27:** 1605-1619

Nesvizhskii AI (2012) Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. Proteomics **12:** 1639-1655

Nikolau BJ, Ohlrogge JB, Wurtele ES (2003) Plant biotin-containing carboxylases. Arch Biochem Biophys **414:** 211-222

Nikolau BJ, Wurtele ES, Stumpf PK (1985) Use of streptavidin to detect biotin-containing proteins in plants. Anal Biochem **149:** 448-453

Ohad N, Shichrur K, Yalovsky S (2007) The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. Plant Physiol **145:** 1090-1099

Ohad N, Yalovsky S (2010) Utilizing bimolecular fluorescence complementation (BiFC) to assay protein-protein interaction in plants. Methods Mol Biol **655:** 347-358

Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res **42:** D358-363

Ozawa T, Kaihara A, Sato M, Tachihara K, Umezawa Y (2001) Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. Anal Chem **73:** 2516-2521

Pauwels L, Ritter A, Goossens J, Durand AN, Liu H, Gu Y, Geerinck J, Boter M, Vanden Bossche R, De Clercq R, Van Leene J, Gevaert K, De Jaeger G, Solano R, Stone S, Innes RW, Callis J, Goossens

A (2015) The RING E3 Ligase KEEP ON GOING Modulates JASMONATE ZIM-DOMAIN12 Stability. Plant Physiol **169:** 1405-1417

Periasamy A, Vogel SS, Clegg RM (2012) FRET 65: a celebration of Forster. J Biomed Opt **17:** 011001

Pertl-Obermeyer H, Schulze WX, Obermeyer G (2014) In vivo cross-linking combined with mass spectrometry analysis reveals receptor-like kinases and Ca(2+) signalling proteins as putative interaction partners of pollen plasma membrane H(+) ATPases. J Proteomics **108:** 17-29

Petrascheck M, Castagna F, Barberis A (2001) Two-hybrid selection assay to identify proteins interacting with polymerase II transcription factors and regulators. Biotechniques **30:** 296-298, 300, 302

Pfleger KD, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). Nat Methods **3:** 165-174

Puchta H (2017) Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr Opin Plant Biol **36:** 1-8

Rahim G, Bischof S, Kessler F, Agne B (2009) In vivo interaction between atToc33 and atToc159 GTP-binding domains demonstrated in a plant split-ubiquitin system. J Exp Bot **60:** 257-267

Rao VS, Srinivas K, Sujini GN, Kumar GN (2014) Protein-protein interaction detection: methods and analysis. Int J Proteomics **2014:** 147648

Rees JS, Li XW, Perrett S, Lilley KS, Jackson AP (2015) Protein Neighbors and Proximity Proteomics. Mol Cell Proteomics **14:** 2848-2856

Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science **339:** 1328-1331

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol **17:** 1030-1032

Rohila JS, Chen M, Cerny R, Fromm ME (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J **38:** 172-181

Roux KJ (2013) Marked by association: techniques for proximity-dependent labeling of proteins in eukaryotic cells. Cell Mol Life Sci **70:** 3657-3664

Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol **196:** 801-810

Serebriiskii I, Estojak J, Berman M, Golemis EA (2000) Approaches to detecting false positives in yeast two-hybrid systems. Biotechniques **28:** 328-330, 332-326

Shoemaker BA, Panchenko AR (2007) Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol **3:** e42

Shoemaker BA, Panchenko AR (2007) Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol **3:** e43

Singh DK, Calvino M, Brauer EK, Fernandez-Pozo N, Strickler S, Yalamanchili R, Suzuki H, Aoki K, Shibata D, Stratmann JW, Popescu GV, Mueller LA, Popescu SC (2014) The tomato kinome and the tomato kinase library ORFeome: novel resources for the study of kinases and signal transduction in tomato and solanaceae species. Mol Plant Microbe Interact **27:** 7-17

Smaczniak C, Li N, Boeren S, America T, van Dongen W, Goerdayal SS, de Vries S, Angenent GC, Kaufmann K (2012) Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. Nat Protoc **7:** 2144-2158

Smart SK, Mackintosh SG, Edmondson RD, Taverna SD, Tackett AJ (2009) Mapping the local protein interactome of the NuA3 histone acetyltransferase. Protein Sci **18:** 1987-1997

Stagljar I, Korostensky C, Johnsson N, te Heesen S (1998) A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc Natl Acad Sci U S A **95:** 5187-5192

Stefan E, Aquin S, Berger N, Landry CR, Nyfeler B, Bouvier M, Michnick SW (2007) Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. Proc Natl Acad Sci U S A **104:** 16916-16921

Stingl K, Schauer K, Ecobichon C, Labigne A, Lenormand P, Rousselle JC, Namane A, de Reuse H (2008) In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification. Mol Cell Proteomics **7:** 2429-2441

Stynen B, Tournu H, Tavernier J, Van Dijck P (2012) Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. Microbiol Mol Biol Rev **76:** 331-382

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res **47:** D607-D613

Tae-Wuk Kim CHP, Chuan-Chih Hsu, Jia-Ying Zhu, Yuchun Hsiao, Tess Branon, Shou-Ling Xu, Alice Y Ting, Zhi-Yong Wang (2019) Application of TurboID-mediated proximity labeling for mapping a GSK3 kinase signaling network in Arabidopsis. bioRxiv

Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol Cell Proteomics **5:** 737-748

Trigg SA, Garza RM, MacWilliams A, Nery JR, Bartlett A, Castanon R, Goubil A, Feeney J, O'Malley R, Huang SC, Zhang ZZ, Galli M, Ecker JR (2017) CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. Nat Methods **14:** 819-825

Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert FM, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H, Lamond A (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol **183:** 223-239

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, Martens L, Witters E, Gevaert K, De Jaeger G (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc **10:** 169-187

Van Leene J, Eeckhout D, Persiau G, Van De Slijke E, Geerinck J, Van Isterdael G, Witters E, De Jaeger G (2011) Isolation of transcription factor complexes from Arabidopsis cell suspension cultures by tandem affinity purification. Methods Mol Biol **754:** 195-218

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, De Bodt S, Maere S, Laukens K, Pharazyn A, Ferreira PC, Eloy N, Renne C, Meyer C, Faure JD, Steinbrenner J, Beynon J, Larkin JC, Van de Peer Y, Hilson P, Kuiper M, De Veylder L, Van Onckelen H, Inze D, Witters E, De Jaeger G (2010) Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana. Mol Syst Biol **6:** 397

Van Leene J, Witters E, Inze D, De Jaeger G (2008) Boosting tandem affinity purification of plant protein complexes. Trends Plant Sci **13:** 517-520

Varshavsky A (1996) The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci U S A **93:** 12142-12149

Villalobos V, Naik S, Piwnica-Worms D (2008) Detection of protein-protein interactions in live cells and animals with split firefly luciferase protein fragment complementation. Methods Mol Biol **439:** 339-352

Vinayagam A, Stelzl U, Wanker EE (2010) Repeated two-hybrid screening detects transient protein-protein interactions. Theoretical Chemistry Accounts **125:** 613-619

Wang Y, Hou Y, Gu H, Kang D, Chen Z, Liu J, Qu LJ (2012) The Arabidopsis APC4 subunit of the anaphase-promoting complex/cyclosome (APC/C) is critical for both female gametogenesis and embryogenesis. Plant J **69:** 227-240

Wang Y, Hou Y, Gu H, Kang D, Chen ZL, Liu J, Qu LJ (2013) The Arabidopsis anaphase-promoting complex/cyclosome subunit 1 is critical for both female gametogenesis and embryogenesis(F). J Integr Plant Biol **55:** 64-74

Xie Q, Soutto M, Xu X, Zhang Y, Johnson CH (2011) Bioluminescence resonance energy transfer (BRET) imaging in plant seedlings and mammalian cells. Methods Mol Biol **680:** 3-28

Yamano H (2019) APC/C: current understanding and future perspectives. F1000Res **8**

Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrzikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M (2011) Next-generation sequencing to generate interactome datasets. Nat Methods **8:** 478-480

Zempleni J, Wijeratne SS, Hassan YI (2009) Biotin. Biofactors **35:** 36-46

Zhong J, Haynes PA, Zhang S, Yang X, Andon NL, Eckert D, Yates JR, 3rd, Wang X, Budworth P (2003) Development of a system for the study of protein-protein interactions in planta: characterization of a TATA-box binding protein complex in Oryza sativa. J Proteome Res **2:** 514-522

Zhou Z, He M, Shah AA, Wan Y (2016) Insights into APC/C: from cellular function to diseases and therapeutics. Cell Div **11:** 9

Scope

Development of a gene-centered method for protein-DNA interaction analysis in plants.

A lot of techniques exist to analyze protein-DNA interactions. As mentioned in chapter one, these techniques can be subdivided in TF-centered and gene-centered methods. Although many TF-centered methods are available, the list of gene-centered tools is less comprehensive and even nonexistent in plants. Identification of protein interactors for a plant specific DNA sequence has to be obtained via a Y1H screen which is accompanied by several drawbacks including the analysis outside the native context. To overcome this problem, a gene-centered *in planta* method is needed. Such a technique would be an important new platform for the elucidation of gene regulatory networks (GRN) in plants. Different *in situ* gene-centered strategies have been reported in other species. One is the ChAP-MS strategy using an exogenous DNA binding sequence and corresponding DNA binding protein, fused to an affinity tag, to target a specific genomic region (Hoshino and Fujii, 2009; Byrum et al., 2012). Implementing crosslinking and subsequent affinity purification results in the specific isolation of the target sequence and its interacting partners. Alternatively, CRISPR can be implemented using a tagged deactivated Cas9 (dCas9) and a locus specific guide RNA (gRNA) (Fujita and Fujii, 2013; Waldrip et al., 2014). Recently, a third option has been reported making use of biotin–based proximity labeling to specifically biotinylate proteins in the neighborhood of a specific genomic region and allowing subsequent stringent affinity purification with streptavidin (Schmidtmann et al., 2016; Myers et al., 2017). The major aim of this doctoral research is to develop a gene-centered method for plant cells, based on the previous mentioned strategies in mammalian cells and yeast. We will test three approaches in *Arabidopsis* cell suspension cultures, being the ChAP-MS strategy (chapter 3), the proximity labeling strategy (chapter 3) and the CRISPR-ChAP-MS strategy (chapter 4). To proof functionality, we analyze the protein interactome of single and multicopy loci for which interacting proteins have been reported.

Identification of APC mitotic substrates.

The list of protein-protein interaction (PPI) techniques is even more elaborate than those for protein-DNA analysis (see chapter 2). However, protein-protein interactions remain a challenging field to investigate due to their different functionalities, dynamics and complexity. Especially the identification of low affinity interactions and dynamic, transient interactions remain to this date a tough task to perform. Different strategies exist to capture transient PPIs of which proximity labeling seems to be the most promising tool, allowing identification of all interacting proteins, including stable,

transient and dynamic interactions (Roux, 2013; Rees et al., 2015). In addition, we recently reported a pull-down protocol (AP-MS) that allowed us to identify the transient interaction between the essential plant growth regulator, the TOR kinase, and its substrates (Van Leene et al., 2019). In a second part of this doctoral research (chapter 5) we will test this pull-down technique for its applicability in discovering mitotic substrates for E3-ligases, here applied on the anaphase promoting complex (APC). Next to that, we will also apply proximity labeling to validate the AP-MS data and to further expand the APC interactome for low affinity interactors.

References

Byrum SD, Raman A, Taverna SD, Tackett AJ (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep **2:** 198-205

Fujita T, Fujii H (2013) Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun **439:** 132-136

Hoshino A, Fujii H (2009) Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. J Biosci Bioeng **108:** 446-449

Myers SA, Wright J, Zhang F, Carr SA (2017) CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus. Molecular & Cellular Proteomics **16:** S63-S63

Rees JS, Li XW, Perrett S, Lilley KS, Jackson AP (2015) Protein Neighbors and Proximity Proteomics. Mol Cell Proteomics **14:** 2848-2856

Roux KJ (2013) Marked by association: techniques for proximity-dependent labeling of proteins in eukaryotic cells. Cell Mol Life Sci **70:** 3657-3664

Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H (2016) Determination of local chromatin composition by CasID. Nucleus **7:** 476-484

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Wahls WP, Taverna SD, Raney KD, Tackett AJ (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics **9:** 1207-1211

Chapter 3

ChAP-MS in plants: an encounter with non-specific interacting proteins.

Caroline Matthijs[1,2], Jelle Van Leene[1,2], Geert Persiau[1,2], Nancy De Winne[1,2], Eveline Van De Slijke[1,2], Dominique Eeckhout[1,2], Francis Impens[3], Geert De Jaeger[1,2]

Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

VIB Proteomics Core, 9000 Ghent, Belgium[3]

Author contributions

CM, JVL and GDJ designed the research; CM, NDW and EVDS performed cloning and transformation experiments; CM and GP performed the research; mass spectrometry was performed by FI and mass spectrometry data was analyzed by DE; CM analyzed the data and wrote the chapter; JVL and GDJ supervised and complemented the writing.

Abstract

Protein-DNA interactions (PDIs) are essential for a proper regulation of the genetic code. Different techniques exist to unravel these PDIs, subdivided in TF-centered and gene-centered methods. To date, an *in planta* gene-centered method is not available, causing Y1H to be one of the only options to analyze the interactome of a specific plant DNA sequence, in addition to some *in vitro* methods. In this chapter we describe two gene-centered applications in *Arabidopsis* cell suspension culture and the corresponding pitfalls we encountered. As a proof of concept we analyzed three different promoters (DR5v2, pETG1 and pCycB1;2) to pull-down well known corresponding transcription factors. We show that prior to pull-down fixating PDIs by cross-linking should be avoided, as this results in the isolation of an excessive amount of non-specific interactions. Although applying a biotin-based proximity labeling strategy, evading the cross-linking step, did not retrieve any known interacting proteins, we conclude that further optimization is needed to obtain an efficient *in planta* gene-centered method based on recently reported gene-centered methods in mammalian cells.

Introduction

A precise regulation of the genetic information encoded in DNA is of utmost importance for the proper functioning of an organism. An appropriate transcription, which can be time dependent, stimuli dependent, cell cycle-phase dependent and tissue-specific, results in cellular functions that allow for changes in growth, differentiation, division and responses to environmental impulses. The basis of this important regulation is the sequence-specific binding of proteins on DNA. These protein-DNA interactions (PDIs) are essential and mutations in one of the two interaction partners can lead to severe malfunctioning of the organism. Unraveling these PDIs is therefore of great value to understand gene regulation.

Techniques used for identification of PDIs can be subdivided in two groups, the transcription factor- (TF) centered and the gene-centered ones. TF-centered techniques, like the high-throughput ChIP-seq technology, identify the DNA loci that are bound by proteins, like TF's. Gene-centered techniques, like the well-known yeast one-hybrid (Y1H) screen, work the other way around. They identify the binding proteins for a specific DNA locus. Although many PDI techniques can be applied in many organisms, there is still one major gap for plants. Gene-centered analysis of a plant locus can only be obtained by performing a Y1H screen, outside the native environment and with a higher chance of false positive and negative results. In addition, Y1H screens only identify direct interactors, missing out the protein complexes generally present on chromatin. To obtain a proteome-wide view of the protein complexes bound on specific plant genomic regions, a new *in planta* gene-centered method is needed.

In the past different *in situ* gene-centered methods have been developed in mammalian cells and yeast, allowing for the identification of proteins present at a defined, small region of chromatin. One tool uses a DNA probe to isolate specific sequences *in vivo* and is called proteomics of isolated chromatin segments (PICh) (Dejardin and Kingston, 2009). This specific DNA probe is linked to an affinity handle making it possible to isolate the target sequence and its interactors via affinity purification. Successful isolation of human telomere sequences and Drosophila Telomere-Associated Sequence (TAS) repeats were obtained, resulting in the identification of well-known and new interactors with mass spectrometry (MS) (Dejardin and Kingston, 2009; Antao et al., 2012; Bartocci et al., 2014). Another approach for the specific isolation of a genomic locus *in vivo* makes use of genetic engineering. A repeat of an exogenous DNA binding sequence is inserted in a cell line in proximity of the genomic region of interest. Furthermore, the complementary exogenous DNA binding protein (DBP) is expressed in the cell and is fused to an affinity handle. After *in vivo*

cross-linking and fragmentation of the chromatin, affinity purification is performed resulting in the identification of interactors with quantitative MS (figure 1A). One technique that makes use of this strategy is insertional chromatin immunoprecipitation (iChIP) (Hoshino and Fujii, 2009; Fujita and Fujii, 2013). Here successful isolation of the IRF-1 promoter (Hoshino and Fujii, 2009), the chicken insulator HS4 (cHS4) (Fujita and Fujii, 2011) and Pax5 1A promoter (Fujita et al., 2015) was obtained in mammalian cells. Also in yeast a similar approach was applied, called chromatin affinity purification with mass spectrometry (ChAP-MS), for the isolation of the GAL1 promoter (Byrum et al., 2012) and to elucidate the double stranded break (DSB) repair mechanism at the MAT locus (Wang et al., 2017).

With the discovery of CRISPR, a new and rapidly developing approach for gene-centered analysis is rising. Successful gene-centered experiments have been performed with CRISPR (Fujita and Fujii, 2013; Waldrip et al., 2014) and new variants are being developed which are making use of the high affinity between biotin and streptavidin. One can fuse a deactivated Cas9 (dCas9) to biotin and apply a stringent purification (CAPTURE) (Liu et al., 2017; Liu et al., 2018) (figure 1B) or one can fuse dCas9 to a promiscuous biotin ligase (BirA*) or an engineered ascorbate peroxidase (APEX2) (figure 1C) allowing for proximity-dependent biotin identification (CasID or C-BERST) (Schmidtmann et al., 2016; Myers et al., 2017; Gao et al., 2018; Myers et al., 2018; Gao et al., 2019; Qiu et al., 2019). This latter approach has been on the rise in recent years and will likely be the method of choice in the future to analyze not only protein-DNA interactions but also protein-protein interactions. The principle of the proximity-based labeling is as follows. To target a specific genomic locus, a dCas9 is expressed together with a locus specific guide RNA (gRNA). To be able to label the proteins in the neighborhood with biotin, dCas9 is fused to BirA* or APEX2. As an alternative the gRNA can be modified with MS2 RNA elements that are bound by a transgenic fusion protein containing the RNA binding protein MS2 and APEX2 (Qiu et al., 2019). To initiate biotinylation, transgenic cells are treated with biotin (in case of BirA*) or biotin-phenol and H_2O_2 (in case of APEX2). Subsequently, biotinylated proteins are pulled-down with a streptavidin resin and are identified by mass spectrometry. Through this way, a new putative transcription factor, ZNF512, was identified at chromocenters in mouse myoblast cells (Schmidtmann et al., 2016). Proximity labeling also confirmed the association of the shelterin complex at telomeres in mammalian cells (Schmidtmann et al., 2016; Gao et al., 2018; Qiu et al., 2019). In addition, single locus genes have been analyzed, revealing known and unknown interactors of the hTERT and c-MYC promoters (Myers et al., 2018).

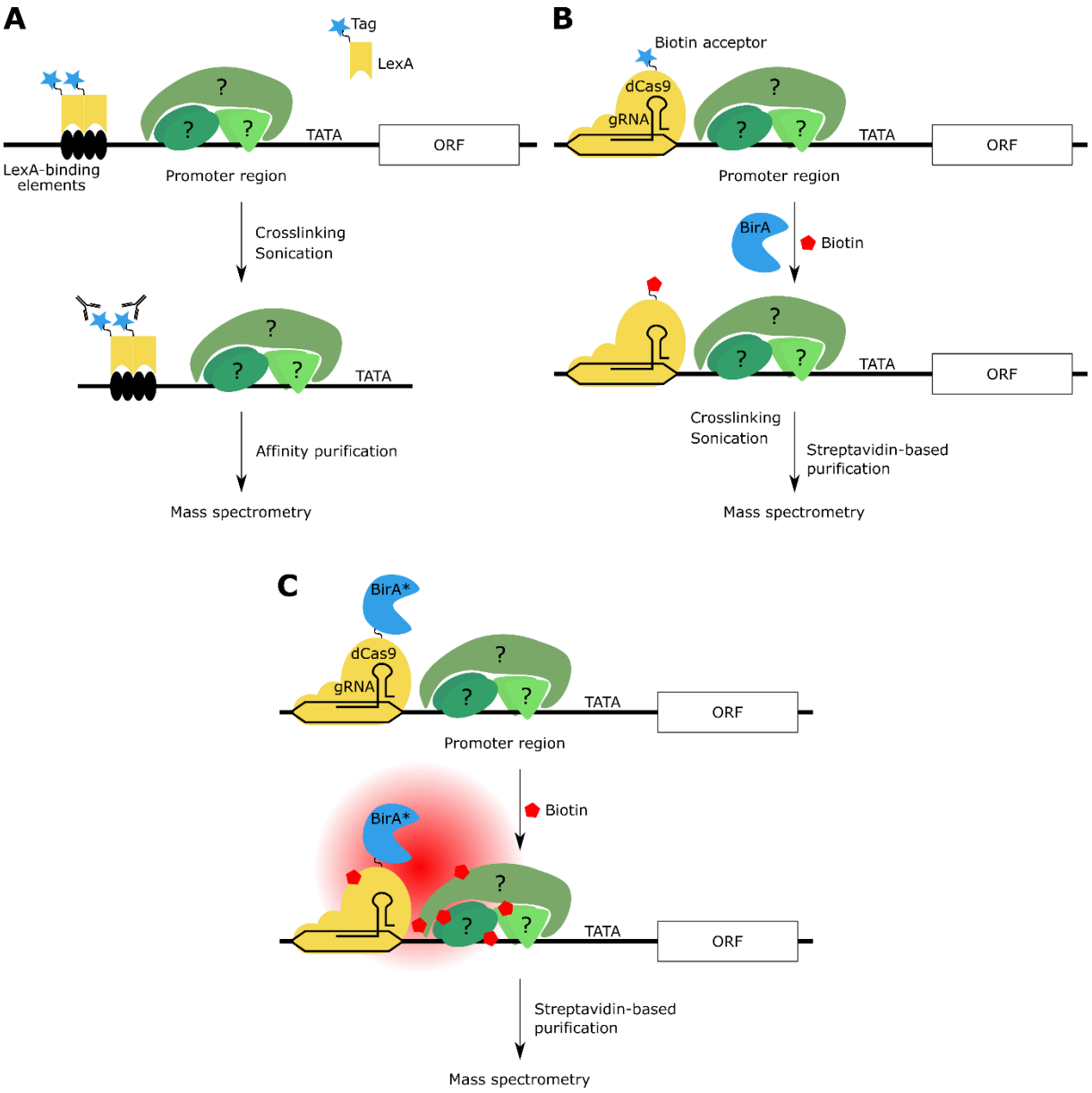


Figure 1: Scheme of different gene-centered methods for protein-DNA interaction analysis at a specific genomic locus. **A)** Representation of the strategy used in insertional chromatin immunoprecipitation (iChIP) and chromatin affinity purification with mass spectrometry (ChAP-MS) (Hoshino and Fujii, 2009; Byrum et al., 2012). LexA-binding elements are inserted in close proximity of the promoter region of interest. Transgenic LexA, fused to an affinity handle (tag), will bind at the specific locus. After fixating all interactions by cross-linking and fragmentation of the chromatin by sonication, affinity purification is performed and isolated proteins can be identified by mass spectrometry. **B)** Representation of the CAPTURE technique (Liu et al., 2018), making use of a deactivated Cas9 (dCas9), fused to a biotin acceptor, and a locus specific gRNA to target a specific genomic region. Induced expression of BirA results in biotinylation of dCas9. Subsequent cross-linking, sonication and streptavidin-based purification allows for isolation of the specific genomic region and its interacting proteins. **C)** Representation of the proximity-dependent biotin identification of the protein interactome for a specific genomic locus (Schmidtmann et al., 2016; Gao et al., 2018). A BirA*-fused dCas9 at the genomic region of interest allows for biotinylation of the neighboring proteins upon biotin addition. Biotinylated proteins are pulled-down with a streptavidin resin and identified with mass spectrometry.

Although, PICh has been applied once in barley for the analysis of centromeric chromatin (Zeng and Jiang, 2016), no other studies have reported the development or application of *in planta* gene-centered methods. There is need for the development of a new, plant-specific method that can yield enough material for the identification of protein-DNA interactions via a quantitative MS analysis. Such a technique would be an important new platform for the elucidation of gene regulatory networks (GRN) in plant cells.

Here we tested new gene-centered *in planta* tools in analogy with ChAP-MS and BirA* proximity-based biotin labeling. ChAP-MS performed on *Arabidopsis* cell suspension culture, allowed us to isolate different DNA sequence of interest but specific identification of bound proteins via quantitative mass spectrometry remains a difficulty. We identified the cross-linking step as the culprit, leading to pull-down of a massif amount of non-specific interacting proteins. Combining ChAP-MS with proximity-based biotin labeling avoids this cross-linking step, but due to the highly active biotin ligase, further optimization is needed to allow for identification of a locus-specific interactome.

Results

ChAP-MS step by step

The scheme of our plant specific ChAP-MS methods is represented in figure 2 and is as follows:

i) A T-DNA construct is generated via multisite gateway cloning (Karimi et al., 2002) containing two cassettes (figure 2A). One cassette contains the DNA sequence of interest, in this case a promoter sequence, flanked by an octarepeat of the prokaryotic lactose operator (LacO) sequence and the β-glucuronidase (GUS) open reading frame (ORF). The second cassette allows constitutive expression of the prokaryotic lac repressor (LacI) fused to the tandem affinity purification (TAP) tag GS^{rhino} (Van Leene et al., 2015) which consists of a protein G tag (ProtG) and a streptavidin-binding peptide (SBP) separated by a very specific rhinovirus 3C protease cleavage site. A kanamycin resistance (Km^R) marker allows for selection of transformed *Arabidopsis* cells.

ii) A transgenic *Arabidopsis* cell suspension culture (in PSB-D, grown in constant dark) is obtained by co-cultivation with *Agrobacteria* containing the T-DNA vector. GS^{rhino}-LacI proteins will interact with the corresponding LacO sequences, while native proteins will bind on the DNA sequence of interest (figure 2B).

iii) To fixate all these interactions, cells are chemically cross-linked *in vivo* with formaldehyde.

iv) After lysis of the cells, chromatin is fragmented via sonication in fragments ranging the size of the LacO repeat and the region of interest.

v) The GS^{rhino}-LacI-DNA complexes are immunoprecipitated with home-made magnetic IgG beads.

vi) Subsequent reverse cross-linking or on-bead digestion with trypsin allows for identification and characterization of the interacting proteins, isolated with the specific target sequence, via mass spectrometry.

Besides promoters, other target sequences can be analyzed with the ChAP-MS technique, like enhancer sequences or specific regulatory DNA motifs which can be put in a repetition. However, specific PDIs are low abundant in a cell and one-step affinity purifications generally yield a big pool of specific and non-specific isolated proteins. Therefore, a label free quantitative MS analysis is recommended for determining which co-enriched proteins are specifically or non-specifically associated with the target sequence. Non-specific proteins can be identified by analyzing a cell

culture only containing the GS^{rhino}-LacI cassette, or by analyzing another random target sequence.

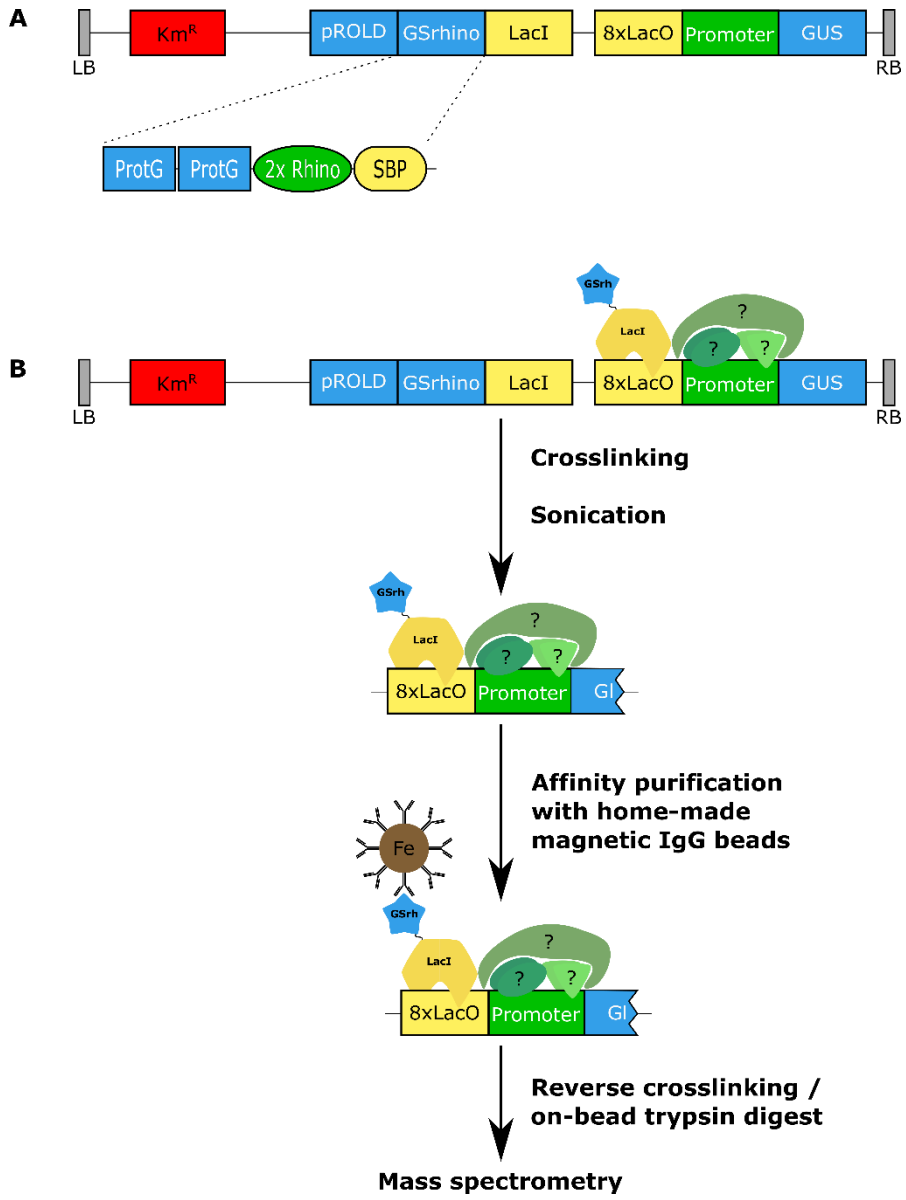


Figure 2: Overview of chromatin affinity purification with mass spectrometry (ChAP-MS). **A)** A T-DNA is composed containing the DNA sequence of interest (e.g. promoter) flanked by an 8 times LacO repeat and GUS ORF, and expresses LacI fused to the GS^{rhino} affinity handle. The GS^{rhino}-tag consists of two protein G's (ProtG) and a streptavidin-binding peptide (SBP) separated by two rhinovirus 3C protease cleavage sites (2xRhino). Plant cells transformed with this T-DNA can be selected by kanamycin resistance (Km^R). **B)** *In vivo* cross-linking fixates all protein-DNA and protein-protein interactions including the binding of GS^{rhino}-LacI on LacO. After cell lysis, chromatin is fragmented by sonication, and the protein-DNA complex of interest is purified by using home-made magnetic IgG beads. After reverse cross-linking or on-bead trypsin digestion, co-enriched proteins can be identified by mass spectrometry.

Cross-linking: a crucial step.

Fixation of the *in vivo* interactions is of great importance when performing a gene-centered analysis and should be optimized to be successful. Cross-linking conditions that are too harsh, will lead to inefficient reverse cross-linking and elution of proteins, while soft cross-linking condition will lead to the loss of interactors during the different affinity purification steps. The most commonly used chemical crosslinker is formaldehyde, which fixates all macromolecular interactions that are in close proximity (± 2 Å) (Hoffman et al., 2015). Formaldehyde cross-linking of our *Arabidopsis* cell suspension culture PSB-D was already optimized for the tandem chromatin affinity purification technique (TChAP) allowing successful isolation of TF specific target sequences (Verkest et al., 2014). Here 0,75% formaldehyde is added to the cell culture for 10 minutes and is subsequently quenched by adding an excess of glycine.

For our ChAP-MS protocol we compared two different cross-linking conditions for their DNA purification tolerance with a non-cross-linking condition. One cell culture was cross-linked with the standard TChAP condition, another with 0,5% formaldehyde for 5 min, and one without cross-linking. Subsequent DNA purification was performed with or without a preceding de-cross-linking step. Cross-linking that is too harsh, will result in lower DNA yield after de-cross-linking, while too soft cross-linking will result in DNA purification even without de-cross-linking (Haring et al., 2007). An optimal cross-linking condition will result in absence of DNA yield without de-cross-linking and efficient DNA purification with de-cross-linking. Cell cultures cross-linked with 0,75% formaldehyde showed low DNA yield without de-cross-linking and after de-cross-linking (figure 3). A concentration of 0,5% formaldehyde still blocks DNA purification without de-cross-linking, though a substantial amount of DNA could be isolated after de-cross-linking. Compared to the DNA yields obtained with 0,5% formaldehyde cross-linking, the standard TChAP cross-linking condition could be too harsh for upcoming experiments. However, because of the successful TChAP results, both cross-linking conditions were further tested for protein fixation and ChAP-MS purification.

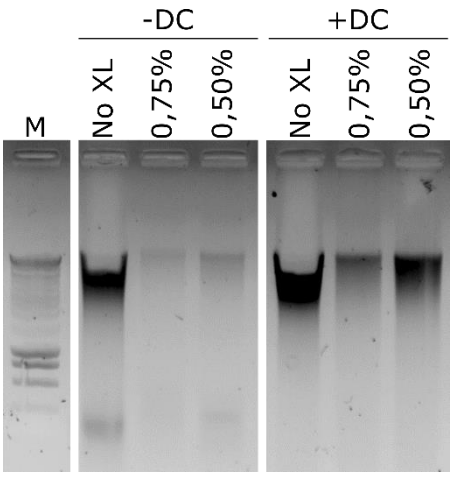


Figure 3: Cross-linking efficiency analysis. Arabidopsis cell cultures were cross-linked with increasing amounts of formaldehyde. Samples were de-cross-linked (+DC) or not (-DC), and DNA was purified. While DNA is efficiently isolated from samples that were not cross-linked (No XL), de-cross-linking is required for the isolation of DNA from cross-linked samples. More DNA is isolated with 0,50% formaldehyde. M: SmartLadder SF (Eurogentec).

The synthetic DR5v2 promoter as a case study.

To proof functionality of the ChAP-MS T-DNA in our *Arabidopsis* cell suspension culture, the DR5v2 promoter was analyzed. The synthetic DR5v2 promoter consists of a repeat of nine auxin response elements (AuxRE) (TGTCGG) which are recognized and bound with high affinity by auxin response factors (ARFs) (Liao et al., 2015). When auxin levels are low, these ARFs are inhibited by auxin/indole3-acetic acid proteins (Aux/IAAs). Upon auxin increase, Aux/IAAs are ubiquitinated and degraded, releasing the ARFs from inhibition and allowing activation or repression of auxin responsive genes and the DR5v2 promoter (Wang and Estelle, 2014). In *Arabidopsis*, there are 23 ARFs of which five (ARF5, ARF6, ARF7, ARF8 and ARF19) are transcriptional activators (Guilfoyle and Hagen, 2007; Li et al., 2016). Of these 23 ARFs, 13 are expressed in our PSB-D cell suspension culture, including the five transcriptional activators of which ARF5 is being the most strongly expressed (unpublished data).

A T-DNA containing the DR5v2 promoter was developed and transformed in our PSB-D culture. In addition, the T-DNA harbored an extra expression cassette, ensuring overexpression of ARF7 fused to a three times human influenza hemagglutinin (3xHA) tag (figure 4A). In this way, a fast evaluation of the T-DNA is guaranteed, allowing ChIP-qPCR and western blot analysis of ARF7. Accumulation of GS^{rhino}-LacI and ARF7-3xHA was observed on western blot (figure 4B), whereas ChIP-qPCR analysis of ARF7-3xHA resulted in an efficient enrichment of the DR5v2 promoter (figure 4C), indicating that the T-DNA construct is functional in our cell culture and allows for binding of ARF TFs on the DR5v2 promoter. Based on the ChIP-qPCR results, cross-linking of the cell culture with 0,5% formaldehyde resulted in a similar fixation of the TF with its target sequence as with 0,75% formaldehyde.

Next, we assessed the functionality of our ChAP-MS protocol for the DR5v2 promoter. After pull-down with magnetic IgG beads, a fraction of the sample was used for DNA purification and qPCR analysis. An efficient enrichment of the DR5v2 promoter was observed after pull-down (figure 4D), indicating that GS^{rhino}-LacI efficiently binds the LacO sequences in the neighborhood of the DR5v2 promoter and affinity purification of GS^{rhino} with the home-made magnetic IgG beads is effective under both cross-linking conditions. Western blot analysis of the remaining pull-down samples revealed clear presence of ARF7-3xHA in the final eluates (figure 4E). Further evidence for the successful application of our ChAP-MS protocol on the DR5v2 promoter was obtained by MS analysis, resulting in the identification of five ARFs (ARF2, ARF5, ARF7, ARF8 and ARF17). On top of that, TOPLESS, TOPLESS-related proteins and histone deacetylases (HDA) were identified, which are known to be recruited by the Aux/IAA inhibitors to auxin responsive genes (table 1) (Wang and Estelle, 2014). However,

Aux/IAA proteins were not detected by MS analysis. Direct interactions of TOPLESS proteins with ARF2 and ARF17 have been observed in a yeast two-hybrid screen (Causier et al., 2012) and could explain their presence in the final eluate.

These results affirm that with our ChAP-MS protocol it is possible to isolate and identify direct and indirect interactors for a specific DNA sequence in our PSB-D culture. However, different MS results were obtained for both cross-linking conditions. While 0,75% formaldehyde resulted in the identification of the five ARFs by unique peptides; 0,5% formaldehyde allowed detection of unique peptide for only three of them (table 1), concluding that although standard TChAP cross-linking results in less efficient DNA purification, this has no effect on the ChAP-MS purification and detection of isolated proteins with MS.

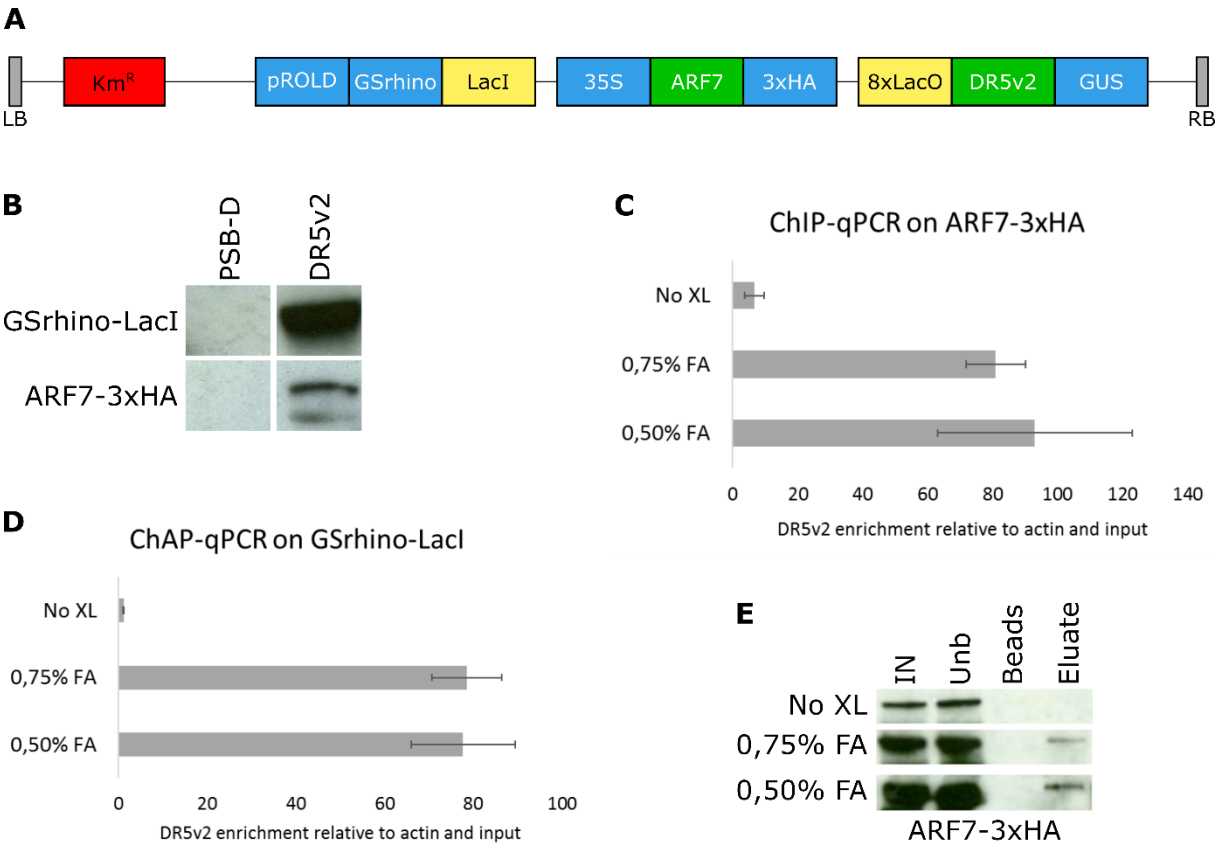


Figure 4: A functional ChAP-MS T-DNA construct for the DR5v2 promoter. **A)** T-DNA construct overexpressing GS^{rhino}-LacI and ARF7-3xHA, including the target promoter DR5v2 flanked by 8 repeats of LacO and GUS open reading frame. **B)** Immunoblot analysis of wild type (PSB-D) and T-DNA transformed cell culture (DR5v2). Accumulation of GS^{rhino}-LacI (59,17 kDa) visualized by Peroxidase Anti-Peroxidase Soluble Complex antibody. Accumulation of ARF7-3xHA (132,68 kDa) visualized by Anti-HA (12CA5). **C-D)** qPCR result for DR5v2 enrichment after pull-down of ARF7-3xHA (C) or GS^{rhino}-LacI (D) under different formaldehyde (FA) cross-linking conditions. **E)** Immunoblot analysis of ARF7-3xHA during pull-down with GS^{rhino}-LacI under different cross-linking conditions. Input samples (IN, 60 µg total protein), unbound fraction (Unb, same volume as IN), fraction of beads after elution (1/60 of total beads) and final eluate (1/60 of total eluate) were loaded and detection was performed using Anti-HA (12CA5). No XL: without cross-linking.

Table 1: Proteins identified by liquid chromatography-tandem MS (LC-MS/MS) after ChAP-MS on the DR5v2 promoter. Number of peptides (# peptides) or number of unique peptides (# unique peptides) identified for the most relevant proteins co-purifying with the bait promoter are shown under the different formaldehyde (FA) cross-linking conditions. No XL: without cross-linking.

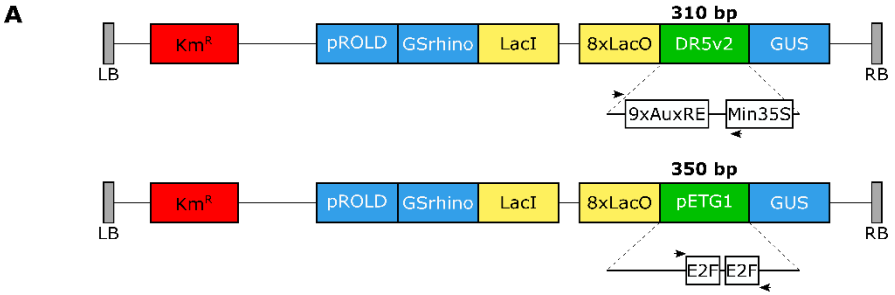
		No XL		0,50% FA		0,75% FA	
Accession number	Protein name	# peptides	# unique peptides	# peptides	# unique peptides	# peptides	# unique peptides
AT5G20730	ARF7	0	0	2	1	2	1
AT5G62000	ARF2	0	0	4	4	2	2
AT1G19850	ARF5	0	0	1	0	2	1
AT5G37020	ARF8	0	0	1	0	2	1
AT1G77850	ARF17	0	0	1	1	1	1
AT1G15750	TPL | TOPLESS	0	0	26	8	24	7
AT1G80490	TPR1 | TOPLESS-related 1	0	0	20	1	19	2
AT3G16830	TPR2 | TOPLESS-related 2	0	0	24	16	23	16
AT5G27030	TPR3 | TOPLESS-related 3	0	0	24	14	24	16
AT3G15880	TPR4 | TOPLESS-related 4	0	0	18	10	16	8
AT4G38130	HD1 | histone deacetylase 1	0	0	7	7	5	5
AT5G22650	HD2B | histone deacetylase 2B	0	0	5	5	3	3
AT3G44750	HDA3 | histone deacetylase 3	0	0	3	3	1	1
AT5G63110	HDA6 | histone deacetylase 6	0	0	2	2	2	2
AT2G27840	HDA13 | histone deacetylase 13	0	0	1	1	1	1
Extra identified proteins		+ 919	+ 824	+ 1926	+ 1888	+1882	+ 1843

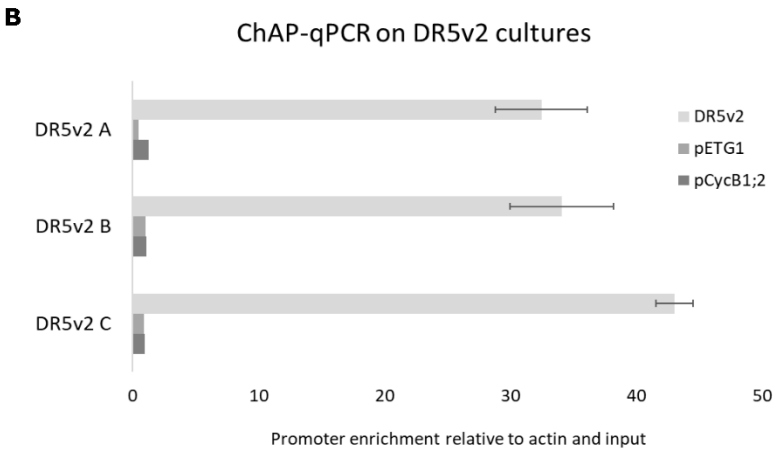
Label Free Quantitative MS analysis uncovers obstacles.

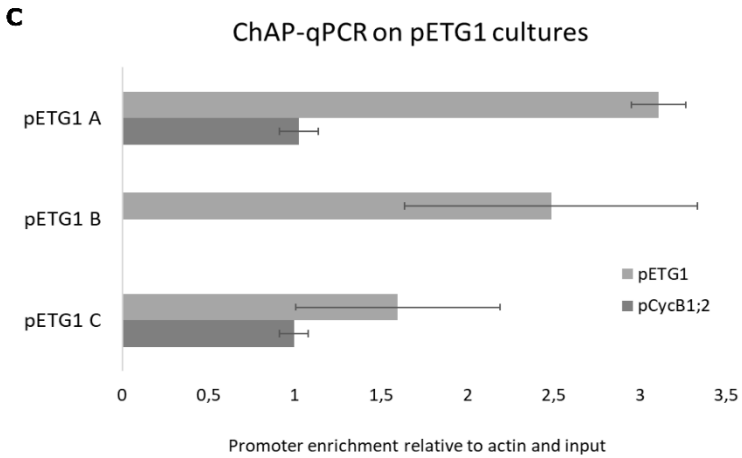
The list of isolated proteins was extensive in the initial experiment. After DR5v2 pull-down 2233 proteins were identified (table 1), many of them being non-specific proteins and common chromatin related proteins. Identifying unknown interactors of a specific DNA sequence becomes impossible when such amount of proteins is identified after a ChAP-MS analysis. To circumvent this problem and facilitate MS analysis, a label free quantitative MS analysis was implemented to identify bona fide interactions.

Different T-DNA constructs were generated allowing the analysis of the DR5v2, ETG1 and CycB1;2 promoters in the absence of an overexpressed complementary TF (figure 5A). For all three promoters transcription regulators have been reported in literature. As mentioned before DR5v2 is bound directly by ARFs and indirectly with AUX/IAA and TOPLESS proteins. Transcription regulation of the E2F TARGET GENE1 (ETG1) by E2F TFs has been previously demonstrated by chromatin immunoprecipitation and tandem chromatin affinity purification (TChAP) (Takahashi et al., 2008; Verkest et al., 2014). And TChAP on the bZIP29 TF revealed strong association with the mitotic B-type cyclin CYCB1;2 promoter near a VIP1 response element (VRE) (Van Leene et al., 2016). Furthermore, RNA-seq data on our PSB-D cell culture revealed that these TFs are efficiently expressed (data not published), which is beneficial for detection of these TFs with MS after ChAP-MS analysis in PSB-D. Two negative controls were included for quantitative analysis, one being the wild type PSB-D culture without T-DNA construct and one cell culture with a T-DNA construct containing the min35S promoter (figure 5A). All cell cultures were cross-linked with 0,75% formaldehyde and for each promoter or control a ChAP-MS analysis was performed in triplicate. Enrichment of the promoter sequences after pull-down was checked with qPCR. Although efficient enrichment of the DR5v2 promoter was obtained, less enrichment was observed for the ETG1 and CycB1;2 promoters (figure 5B-D).

In total 5972 proteins were identified including ARF, E2F and bZIP transcription factors. A quantitative analysis for each promoter versus the other set of promoters and controls (CTset) was executed. However, no significant enrichment of the known transcription factors and co-regulators was observed for one of the promoters under a false discovery rate (FDR) of 0.01 (figure 6). Although some ARFs tend to be more enriched for the DR5v2 promoter (located at the right site of the volcano plot), E2F and bZIP proteins are too (figure 6A). The same also applies to the ETG1, CycB1;2 and min35S promoters (figure 6B-D). In addition, the same transcription factors were also isolated after pull-down in wild type PSB-D culture (data not included), indicating that identification of these TFs is due to non-specific binding onto the affinity resin.







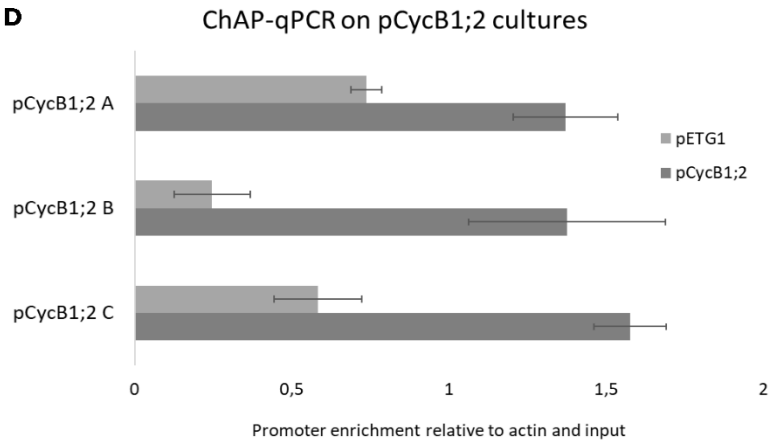


Figure 5: Functional ChAP-MS T-DNA constructs for the DR5v2, ETG1 and CycB1;2 promoters. **A)** T-DNA constructs overexpressing GS^{rhino}-LacI, including the target promoters DR5v2, pETG1 or pCycB1;2, or the negative control Min35S promoter, flanked by 8 repeats of LacO and GUS open reading frame. Length and cis-regulatory elements (AuxRE, E2F and VRE) of the promoters is indicated. Arrows: position of forward and reverse qPCR primers. **B-D)** qPCR results for DR5v2, pETG1 and pCycB1;2 enrichment after triplicate ChAP pull-downs on cell suspension cultures transformed with ChAP-MS T-DNA constructs containing DR5v2 **(B)**, pETG1 **(C)** or CycB1;2 **(D)**.

After re-evaluation, E2Fa was also identified in the initial DR5v2 MS data under both cross-linking conditions (data not shown). Moreover, without cross-linking none of the TFs were isolated (table 1), while with cross-linking of PSB-D they do. This indicates that cross-linking has a major impact on the affinity purification and leads to the non-specific isolation of more than a quarter of all the encoded proteins present in *Arabidopsis*.

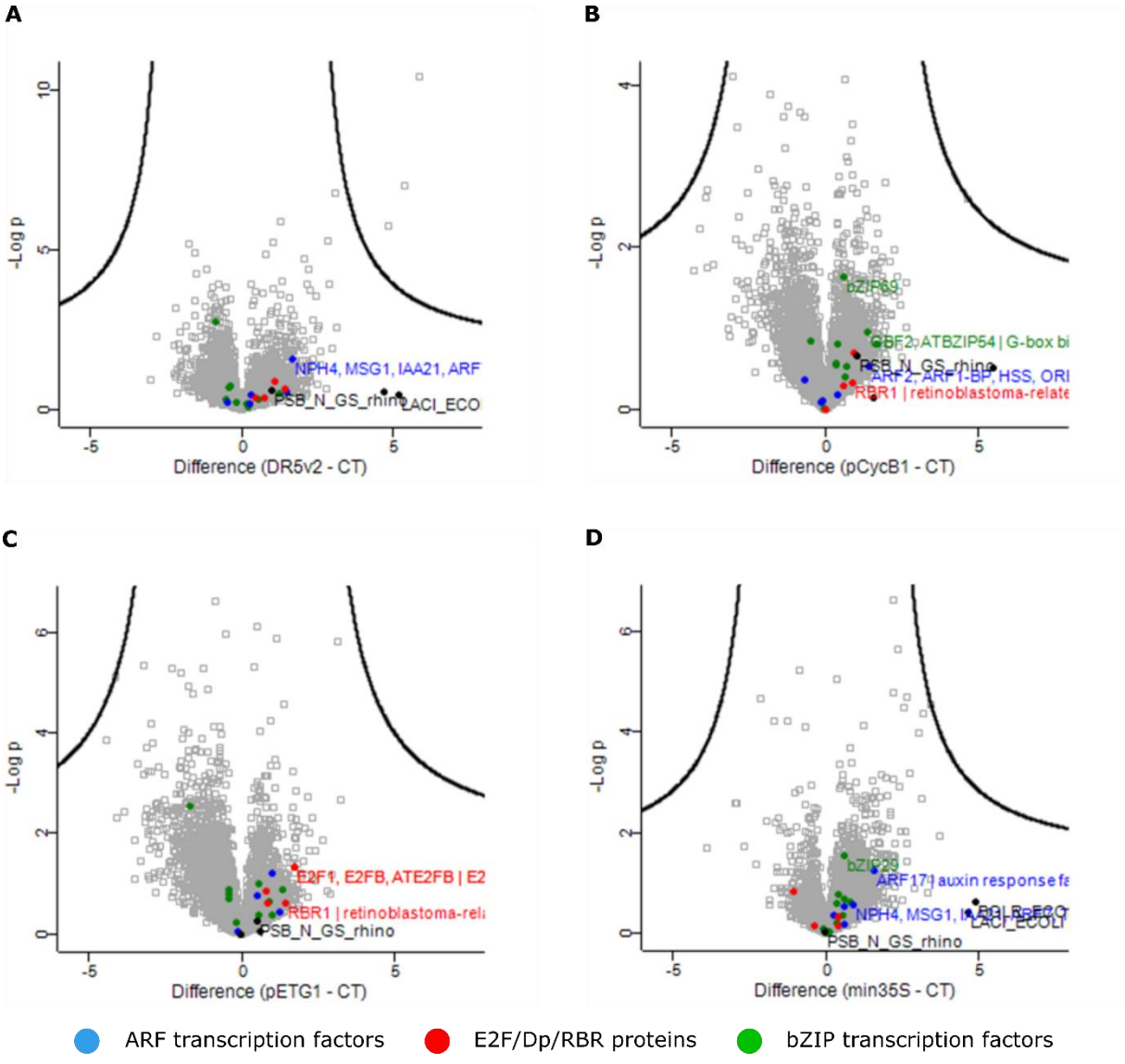


Figure 6: Label free quantitative mass spectrometry analysis on ChAP-MS pull-downs of the DR5v2 (A), CycB1;2 (B), ETG1 (C) and Min35S (D) promoters. A quantitative analysis for each promoter versus the other set of promoters and wild type control (CTset) was executed by comparing MaxQuant LFQ intensity values in Perseus. Proteins were graphed by fold change (Difference) and significance (-Log p) and thresholds were set at a false discovery rate of 0.01 and an S0 of 1. Blue dots represent ARF transcription factors. Red dots represent E2F and DP transcription factors and RBR negative regulator. Green dots represent bZIP transcription factors. Black dots represent LacI and the GS^{rhino} affinity handle.

TurboID-ChAP-MS: integration of proximity labeling

In order to prevent the pull-down of massive amounts of non-specific and common chromatin interacting proteins, we needed to avoid the formaldehyde cross-linking step in our ChAP-MS protocol. Without cross-linking, *in vivo* interactions are no longer fixated and are easily lost after pull-down. Therefore, we still needed to mark interacting proteins for a specific locus *in vivo* before we perform an affinity purification. A gene-centered method that does not use cross-linking for identification of PDIs, is the proximity-dependent biotin identification with BirA* or APEX2 (Schmidtmann et al., 2016; Gao et al., 2018). We implemented this proximity-based labeling in our ChAP-MS protocol by replacing the GS^{rhino} tag with TurboID, a mutated version of the biotin ligase BirA with greater proximity labeling efficiency (Branon et al., 2018) (figure 7A). We also included a 65-amino acid long linker between TurboID and LacI, which increases the labeling radius (Deepanksha Arora, 2019). Hereby, we are able to biotinylate all proteins that come in the neighborhood of the TurboID-LacI, who still resides at the promoter sequence of interest. Subsequent streptavidin based affinity purification can be applied and biotinylated proteins are identified by mass spectrometry (figure 7B).

TurboID-ChAP-MS T-DNA constructs were developed for the DR5v2 and min35S promoter (figure 7A). Accumulation of TurboID-LacI in transformed cell suspension cultures was checked on western blot, showing similar protein levels for both constructs (figure 7C). Accumulation of GUS was also analyzed and we observed a lower level of GUS protein with the min35S promoter, as was expected (figure 7C). Also after qPCR analysis, GUS expression levels were much lower in the cultures transformed with the min35S T-DNA construct (figure 7D). We tried to augment the GUS expression by stimulating the ARF binding on the DR5v2 promoter by first blocking the auxin signal transduction pathway with the auxin antagonist PEO-IAA (Hayashi et al., 2012) for one hour, followed by removal of the inhibitor and addition of synthetic auxin (NAA). Although PEO-IAA treatment resulted in a steep decline of GUS expression, subsequent removal of PEO-IAA and addition of NAA did not enhance GUS expression. On the contrary, expression remained absent for 4 hours after NAA addition (figure 7D).

We performed TurboID-ChAP-MS on the DR5v2 promoter and min35S promoter in triplicate, implementing a label free quantitative MS analysis to specifically enrich for DR5v2 neighboring proteins. In total 2126 proteins were identified of which 85 were specifically enriched with the DR5v2 promoter using a false discovery rate of 0.5 (figure 7E). Although ARFs were co-purified, they were not significantly enriched with the DR5v2 promoter. In addition, bZIP transcription factors and E2Fb were present in the

large MS data set. A closer look at the DR5v2-enriched proteins revealed the presence of 63 proteins which do not localize to the nucleus, making them false positives. The other 22 nuclear located proteins include histone H2A variants, RNA polymerase associated proteins and chromatin remodeling proteins (table 2). Interestingly, the two H2A variants that we retrieved, HTA9 and HTA11, have both been linked with chromatin silencing (Dai et al., 2018), suggesting that our T-DNA construct is in a repressed state. However, based on the expression level of GUS, this seems unlikely.

Next to these housekeeping nuclear proteins, we also identified the transcription repressor JAZ10. JAZ proteins bind and inactivate TFs, whether or not combined with the recruitment of the corepressor TOPLESS (Pauwels and Goossens, 2011). In the presence of the phytohormone jasmonate, the inhibitory effect of JAZ is nullified by SCF^{COI1}-mediated degradation. It has been demonstrated that JAZ proteins interact with a wide range of transcription factors (Pauwels and Goossens, 2011), and methyl jasmonate (MeJA)-dependent upregulation of a DR5 reporter constructs in *Arabidopsis* has been reported (Sun et al., 2009; Hoffmann et al., 2011), suggesting a role for JAZ proteins in regulating auxin responsive genes.

Altogether, our TurboID-ChAP-MS approach did not result in a desirable outcome and is still far from optimal. Also with TurboID-ChAP-MS, we stumble upon the pull-down of a large amount of non-specific biotinylated proteins, which prevent us to identify locus-specific interacting proteins.

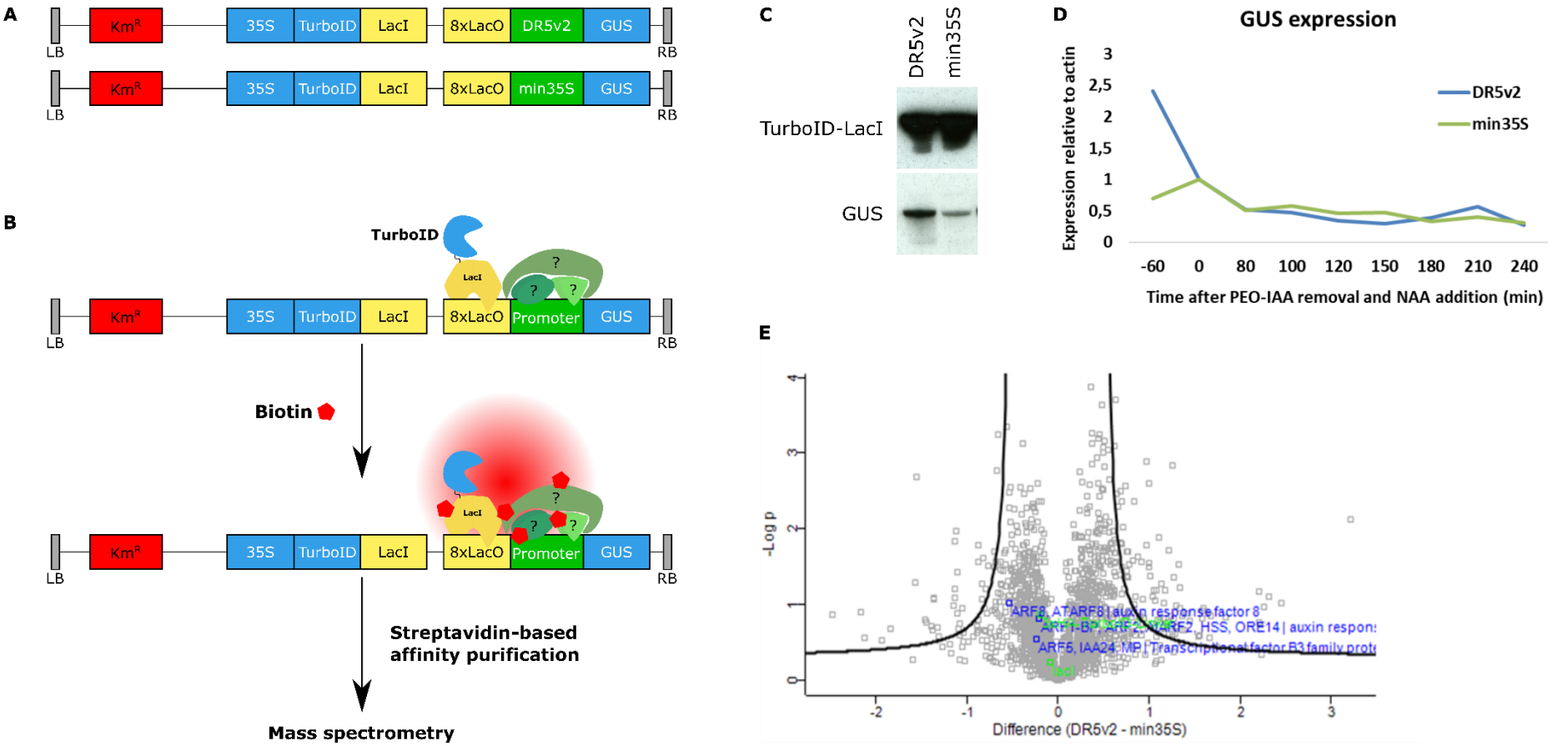


Figure 7: TurboID-ChAP-MS: scheme and expression analysis. **A)** T-DNA constructs overexpressing TurboID-LacI, including the target promoter DR5v2 or the negative control min35S promoter, flanked by 8 repeats of LacO and GUS open reading frame. **B)** Scheme of TurboID-ChAP-MS. TurboID-LacI is expressed and binds LacO sequences upstream of the promoter of interest. Upon the addition of biotin, TurboID will biotinylate the neighboring proteins. Biotinylated proteins are pulled-down with a streptavidin resin and identified with mass spectrometry. **C)** Immunoblot analysis of T-DNA transformed cell cultures with DR5v2 or min35S promoter. Accumulation of TurboID-LacI was visualized by anti-HA antibody. Accumulation of GUS was visualized by anti-GUS antibody. **D)** qPCR result for GUS expression in T-DNA transformed cell cultures with DR5v2 or min35S promoter at different time points before and after PEO-IAA treatment. NAA was added at time point zero. **E)** Volcano plot representing label free quantitative mass spectrometry analysis on triplicate TurboID-ChAP-MS pull-downs of the DR5v2 and Min35S promoters. Significant enriched proteins for the DR5v2 promoter are situated at the right side of the plot. Proteins were graphed by fold change (Difference) and significance (-Log p) and thresholds were set at a false discovery rate of 0.5 and an S0 of 0.8 to select for significantly enriched proteins. Blue dots represent ARF transcription factors. Green dots represent LacI and TurboID.

Table 2: DR5v2 specifically enriched proteins, located in the nucleus, identified by liquid chromatography-tandem MS (LC-MS/MS) after TurboID-ChAP-MS. Fold change (Difference) and significance (-LOG (p-value)) were determined and significantly enriched proteins were selected with a false discovery rate of 0.5 and an S0 of 0.8. Cellular localization was determined by GO enrichment for cellular compartment by PANTHER14.1 (Mi et al., 2013).

Accession number	Protein name	-LOG(P-value)	Difference	Cellular localization
AT5G14600	TRM61, tRNA methyltransferase 61	2,12	3,21	nucleus
AT3G54560	HTA11, histone H2A 11	1,10	2,31	nucleus
AT1G72390	PHL, phytochrome-dependent late-flowering	0,96	2,26	nucleus
AT5G56670	Ribosomal protein S30 family protein	0,75	2,21	nucleus/cytosol
AT1G52740	HTA9, histone H2A protein 9	0,84	1,68	nucleus
AT5G13220	JAZ10, Jasmonate-ZIM-domain protein 10	0,88	1,61	nucleus
AT1G15440	PWP2 , periodic tryptophan protein 2	0,95	1,59	nucleus
AT3G10300	Calcium-binding EF-hand family protein	0,92	1,58	nucleus
AT1G35830	VQ motif-containing protein	0,58	1,56	nucleus
AT3G57660	NRPA1, nuclear RNA polymerase A1	0,58	1,55	nucleus
AT3G52250	POWERDRESS	0,65	1,37	nucleus
AT2G39730	RCA, rubisco activase	1,53	1,32	nucleus/cytosol/plasma membrane/plastid
AT5G60410	SIZ1, small ubiquitin-like modifier (SUMO) E3 ligase	1,36	1,27	nucleus
AT4G21710	NRPB2, subunit of DNA-dependent RNA polymerase II	0,61	1,25	nucleus/cytosol
AT3G06860	MFP2, multifunctional protein 2	1,02	1,16	nucleus/cytosol
AT3G19760	EIF4A-III, eukaryotic initiation factor 4A-III	0,79	1,10	nucleus/cytosol
AT5G67630	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0,74	1,03	nucleus
AT3G15120	BRP1, BRAT1 PARTNER 1	0,85	0,90	nucleus/plasma membrane
AT3G19980	FYPP3, flower-specific, phytochrome-associated protein phosphatase 3	1,10	0,85	nucleus/cytosol
AT5G52470	FBR1, fibrillarin 1	1,01	0,83	nucleus
AT2G16950	TRN1, transportin 1	1,52	0,77	nucleus/cytosol
AT5G09900	RPN5A, regulatory particle non-ATPase subunit 5A	2,03	0,71	nucleus/cytosol

Discussion

ChAP-MS does not identify promoter-specific PDIs.

A precise regulation of the genetic information encoded in DNA is of utmost importance for the proper functioning of an organism. Unraveling the identity and location of transcription regulators is therefore crucial. Although many PDI techniques can be applied for many organisms, there is still one major gap for plants. Identification of upstream regulators of a specific plant gene can only be obtained by performing a Y1H screen, outside the native environment, with a higher chance of false positive and negative results and without the identification of indirect interactors. To obtain a genome-wide view of the protein complexes bound on specific plant genomic regions, gene-centered methods like ChAP-MS (Byrum et al., 2012), iChIP (Hoshino and Fujii, 2009; Fujita and Fujii, 2012) and the proximity-dependent biotin identification techniques (Schmidtmann et al., 2016; Gao et al., 2018; Myers et al., 2018; Qiu et al., 2019) are needed. There is need for the development of a new, plant specific method that can yield enough material for the identification of protein-DNA interactions via a quantitative MS analysis. Such a technique would be an important new platform for the elucidation of gene regulatory networks (GRN) in plant cells.

Here we tried to develop a new gene-centered *in planta* tool by adopting the modus operandi of ChAP-MS and iChIP in our *Arabidopsis* cell suspension culture. A T-DNA construct was created, consisting of the DNA sequence of interest (e.g. a promoter) flanked by the prokaryotic LacO sequence and a cassette allowing expression of GS^{rhino}-tagged prokaryotic LacI. Functionality of this T-DNA in our PSB-D culture was demonstrated by efficient accumulation of GS^{rhino}-LacI and binding of the ARF7 TF on the DR5v2 promoter. Pull-down of GS^{rhino}-LacI also resulted in an effective enrichment of the promoter sequence of interest, illustrating an efficient binding of the lac repressor on LacO in plant cells. Subsequent identification of ARFs and TOPLESS proteins with MS, lead to the presumption that our plant specific ChAP-MS protocol was operational.

However, label free quantitative MS analysis of the DR5v2, ETG1 and CycB1;2 promoters resulted in a less positive observation. Although the promoter sequences were enriched after pull-down, none of the expected upstream regulators were specifically enriched for the three promoters. The corresponding TFs were present in the extensive MS data sets. However, their presence in the negative control sets indicates a non-specific purification of them. In addition, more than a quarter of all encoded proteins present in *Arabidopsis* were identified after pull-down of the wild type PSB-D culture, all being non-specific interactions with the affinity resin. This bottleneck,

co-purification of a high amount of non-specific interactions, prevents us of filtering out the genuine protein interactions of a specific DNA locus.

Comparison of different ChAP-MS techniques.

The plant-specific ChAP-MS tool using the prokaryotic LacI-LacO interaction in combination with formaldehyde cross-linking and GS^{rhino} pull-down has not resulted in the desired outcome. However similar approaches in mammalian cells and yeast were successful (Hoshino and Fujii, 2009; Byrum et al., 2012; Fujita and Fujii, 2012). They used the prokaryotic LexA protein for specific isolation of a DNA locus resulting in approximately a 5-fold enrichment of the target sequence relative to actin (Byrum et al., 2012). However, our pull-down on the DR5v2 promoter yielded a higher enrichment of almost 80-fold relative to actin. This higher enrichment can be explained by the different dissociation constant (K_d) of LexA and LacI for their binding site. While LexA has a K_d of ~10^{-10} M (Lewis et al., 1994; Chen and Bundschuh, 2014), LacI binding to LacO sequences has a stronger affinity with a K_d of ~10^{-13} M (Riggs et al., 1970; Levens and Howley, 1985). This strong interaction also manifests in the fact that cross-linking is not necessary to stabilize the binding of LacI to LacO during ChIP experiments (Newell and Gray, 2010). Another difference between these gene-centered approaches are the affinity tags. While in mammalian cells FLAG is used and in yeast ProtA, we applied the GS^{rhino} TAP tag containing ProtG. All three tags have a similar affinity for their interacting partner, with a K_d of ~10^{-9} M (Wegner et al., 2002; Saha et al., 2003; Yang et al., 2003). Although the FLAG tag and ProtA are smaller than the GS^{rhino} tag, this does not prevent binding of true interactors on the promoter as we see binding of ARF7-3xHA on the DR5v2 promoter. Not finding bona fide interactions in plant cells is not due to the use of LacI, GS^{rhino} or an inefficient purification of the target sequence.

A big difference with the ChAP-MS protocol in yeast, is the amount of starting material. While we start with 20 grams (± 3 x 10^8 cells) of PSB-D cells, they use 5 x 10^{11} of yeast cells. Consequently, this higher amount of cells allows for elevated purification of the low abundant protein-DNA interactions, increasing the final yield of DNA binding proteins and improves the identification of them with MS. Because of the small size of yeast (3-5 µm) a high amount of cells can be obtained quite easily. For big plant cells which have a length of 100-500 µm this is less straightforward. Increasing the starting material to an amount of 10^{11} plant cells is just not feasible. Comparison with the amount of mammalian cells used during iChIP experiments shows that our starting amount is not that low either. iChIP experiments start with an amount of around 4 x 10^7 cells, almost 10 times less.

Another distinction between ChAP-MS in yeast and plants is the wash buffer applied to get rid of non-specifically interacting proteins as much as possible. In yeast a more stringent buffer of 1M NaCl and 1M urea is used, while we applied different washing steps with increasing salt and detergent levels. This stringent washing resulted in the highest enrichment of the GAL1 locus in yeast after pull-down (Byrum et al., 2012). However, implementing this 1M NaCl/1M urea buffer in our ChAP-MS protocol resulted in a huge reduction of target enrichment (data not shown) and was not used any further.

The only other difference that remains is the activation of transcription by induction. Both Hoshino and Fujii (2009) and Byrum et al. (2012) applied stimuli to activate the transcription of their gene of interest. Both observed that without stimuli no TFs and polymerases could be identified after pull-down, while with stimuli transcription regulators were retrieved (Hoshino and Fujii, 2009; Byrum et al., 2012). Induction of transcription results in enrichment of TFs binding on the promoter of interest and as such increase the chance of identifying them with quantitative mass spectrometry. We applied a similar approach to stimulate ARF binding on the DR5v2 promoter by temporarily inhibiting the auxin signal transduction pathway with PEO-IAA and reactivation of the pathway with synthetic auxin (NAA). However, this resulted in no additional expression of GUS, on the contrary, expression remained absent after re-adding NAA. Other strategies should be applied if one wants to stimulate ARF binding on the DR5v2 promoter. For instance, another synthetic auxin could be applied like 2,4-D which shows stronger physiological effects on *A. thaliana* compared to NAA (Trenner et al., 2017). Similarly, for the analysis of the ETG1 promoter, which is activated by E2Fa at onset DNA replication, synchronization of the cell culture at S-phase via sucrose starvation and repletion, could enhance E2Fa binding and increase the chance for identification with MS.

Formaldehyde cross-linking as a culprit.

Although there are some technical differences between the published in-situ gene-centered techniques and our protocol, none of them can explain the negative outcome of our experiments in plant cells. However, our ChAP-MS data revealed that formaldehyde cross-linking of wild type PSB-D culture results in the pull-down of more than a quarter of all encoded proteins present in *Arabidopsis*. As a result, many chromatin-related proteins including transcription factors are co-purified as non-specific interactors of the affinity resin. Due to this, it was impossible to filter out the genuine protein interactions of a specific DNA locus with our ChAP-MS protocol. In order to prevent the pull-down of massive amounts of non-specific and common

chromatin interacting proteins, we avoided the formaldehyde cross-linking step by implementing biotin-based proximity labeling using TurboID (Branon et al., 2018). The use of TurboID comes with various benefits. We not only avoid cross-linking, the high labeling efficiency of TurboID allows for a short biotinylation time span with low concentration of exogenous biotin, even in plants (Mair et al., 2019). In addition, TurboID labeling can be performed under normal plant growth conditions (Mair et al., 2019) and with a smaller amount of starting material (Schmidtmann et al., 2016). Because of the strong affinity between biotin and streptavidin, stringent washing steps can be applied during the pull-down, decreasing the amount of non-specific interactions with the affinity resin. However, it must be taken into account that proteins are labeled, because they are in close proximity with TurboID, and as such may not be true interactors of the specific DNA locus. Subsequent validation with a TF-centered tool, like ChIP-seq, is a necessity.

TurboID-ChAP-MS is highly active.

Biotin-based proximity labeling has been applied for the discovery of DNA locus-specific interactions in mammalian cells by co-expressing a dCas9 fused to a biotin ligase or ascorbate peroxidase with a locus-specific gRNA (Schmidtmann et al., 2016; Myers et al., 2017). We applied a similar approach, replacing the GS^{rhino} with the modified BirA* enzyme, TurboID. However, this also resulted in an extensive list of pulled-down proteins, yet with DR5v2 specifically enriched proteins, but of which many could be discarded as false positives (no nuclear localization) and common chromatin related proteins. Although ARFs were pulled-down, they were not significantly enriched for the DR5v2 promoter, meaning that biotinylation of these TFs also occurred in cells transfected with the min35S T-DNA construct. Our data suggests a non-specific biotinylation of all types of cellular proteins, which can be explained by the overexpression of the TurboID-LacI fusion protein. Only a small portion of the TurboID-LacI proteins can bind the LacO sequences on the T-DNA construct, while the surplus continuously roams in the cell. Upon addition of biotin, not only proteins in close proximity of the DR5v2 or min35S promoter are biotinylated, but also every protein that passes by the labeling radius of free TurboID-LacI, in and outside the nucleus. In addition, we let the biotinylation continue for 24 hours, leading to an increased accumulation of non-specific biotinylated proteins. Furthermore, TurboID is highly active and can use endogenous levels of biotin to label proteins during the entire time-span of cell growth, which only increases the list of false positives.

Lowering the expression level of TurboID-LacI could be an option to improve the TurboID-ChAP-MS approach. This has been done to elucidate the protein interactome

of the FAMA transcription factor in *Arabidopsis* stomatal guard cells, where FAMA-TurboID was under control of the FAMA promoter (Mair et al., 2019). However, they also observed that more than half of the pulled-down proteins with FAMA-TurboID, were also isolated with a nuclear localized TurboID. Because the nucleus is a small subcellular compartment with a dense amount of proteins, non-specific proximity labeling is easily obtained. To identify the genuine FAMA interacting partners, Mair and coworkers implemented negative controls, including a nuclear localized TurboID under control of the FAMA promoter and an untreated (without addition of biotin) FAMA-TurboID sample, allowing them to filter out non-significantly enriched proteins. Alternatively, inducible expressing of TurboID-LacI would decrease the amount of biotin labeled proteins due to endogenous biotin levels. This strategy has been implemented in the gene-centered analysis tools C-BERST and GloPro (Myers et al., 2017; Gao et al., 2018; Myers et al., 2018). In both approaches a dCas9-APEX2 fusion protein is under control of a doxycycline-inducible promoter. Expression of dCas9-APEX2 is activated only 12 to 24 hours before addition of biotin-phenol and H_2O_2. An additional advantage of using APEX2 instead of a biotin ligase, is that endogenous biotin cannot be used by APEX2 for labeling of proteins, avoiding the accumulation of non-specific labeled proteins. Only in presence of biotin-phenol and H_2O_2, proximity labeling is initiated. Strikingly, this proximity labeling only takes one minute in both approaches, which contribute to a lower rate of false positives. Additionally, to avoid labeling of non-nuclear proteins, dCA9-APEX2 is provided with a nuclear localization signal (NLS).

Future perspectives

A gene-centered analysis of PDIs in plants will need further optimization of the TurboID-ChAP-MS approach avoiding as much as possible non-specific labeling. TurboID-LacI protein levels should be kept low and even absent during cell growth, by controlling expression with an inducible promoter. In addition, we should direct TurboID-LacI to the nucleus by implementing a NLS. The time-span of biotin labeling should be kept short, a property for which TurboID has been developed (Branon et al., 2018). TurboID already labels a significant amount of proteins after 10 minutes biotin treatment, however it has been shown that longer treatments result in the identification of more relevant interactions with MS (Mair et al., 2019). To identify the FAMA interactome, 3 hours of biotin treatment was sufficient to identify 47 specific interacting candidate proteins. It would be worthwhile to reduce our 24 hours biotin treatment to 3 hours, to see a decrease in non-specific labeling. Alternatively, we could replace TurboID with APEX2, thereby limiting the biotinylation time-span to one minute.

However, this is accompanied by the addition of expensive biotin-phenol and toxic H_2O_2. Furthermore, plants contain high endogenous peroxidase activity, making APEX2-based proximity labeling unsuitable for plants. Anyhow, proximity labeling studies should always be combined with proper negative controls, to filter out the native biotinylated proteins and the unavoidable non-specific labeling, especially when studying PDI's in a dense nucleus. In analogy with the FAMA-TurboID study, multiple filtering steps should be applied, relative to a wild type sample, a nuclear located single TurboID sample and an untreated TurboID-LacI sample. Of course, combining this TurboID strategy with the CRISPR/dCas9 mediated chromatin purification system, like enChIP and CRISPR-ChAP-MS (Fujita and Fujii, 2013; Waldrip et al., 2014), will allow us to generate even more relevant data by targeting the native locus of interest and could result in an *in planta* gene-centered PDI tool that can be used as an alternative to Y1H analyses.

Materials and methods

Cloning of (TurboID-)ChAP-MS T-DNA constructs

Generation of T-DNA constructs was obtained by following a gene stacking approach with the use of the MultiSite Gateway technology (Karimi et al., 2002). The coding sequence of LacI flanked by AttB2R/AttB3 gateway recombination sites was amplified from the x35eglsp (CmR) vector (Matzke et al., 2003) and subjected to a BP reaction with pDONRP2RP3 to generate the entry vector pEN-R2L3-LacI. In a similar way pEN-L1L2-LacI was generated using pDONR221. The coding sequence of ARF7 flanked by AttB1/AttB2 was amplified from Arabidopsis cDNA, while the promoter sequences pETG1 and pCycB1;2 flanked by attB1/attB2 were amplified from Arabidopsis gDNA, and were subsequently subjected to a BP reaction with pDONR221 to generate the corresponding entry vectors. The entry vectors containing 8 repeats of LacO flanked by AttL4/AttR1, 8xLacO-min35S flanked by AttL1/AttL2 were obtained by DNA synthesis in the pUC57 vector (GenScript). Generation of the GS^{rhino} entry vector is described elsewhere (Van Leene et al., 2015). An entry vector containing 3xHA-TurboID preceded by an omegaleader and followed by an 13xG4Slinker was also obtained by DNA synthesis in the pUC57 vector (GenScript). DR5v2 was amplified from pGIIM/LIC_SwaI-ntdTomato-DR5v2::n3EGFP vector (Liao et al., 2015) flanked by AttB1/AttB2 gateway recombination sites and subjected to a BP reaction in pDONR221 to generate the corresponding entry vector. The p35S, 3xHA and GUS vectors as well as the pK8m43GW2 and pK7m24GW2 destination vectors can be retrieved at https://gateway.psb.ugent.be/. Subsequent MultiSite Gateway recombination in the pK8m43GW2 and pK7m24GW2 destination vectors (Karimi et al., 2002) generated the expression vectors pK8:pROLD-GS^{rhino}-LacI and pK7:p35S-TurboID-LacI, respectively. To insert a second Multisite Gateway cassette, the expression vectors were linearized via AvrII/XbaI restriction followed by ligation of the XbaI restricted fragment from pXb2M43GW7. A second multisite Gateway recombination reaction, combining the promoter sequence of interest with the 8xLacO entry vector and GUS open reading frame (ORF) resulted in the final T-DNA constructs and were transformed to *Agrobacterium tumefaciens* strain C58C1RifR (pMP90) by electroporation. To generate pK8:pROLD-GS^{rhino}-LacI-35S-ARF7-3xHA-8xLacO-DR5v2-GUS, a second Gateway recombination reaction was performed combining the 35S promoter with the ARF7 and 3xHA entry vectors with the pK8:pROLD-GS^{rhino}-LacI-m34GW destination vector, followed by a second linearization step with AvrII to ligate a third Multisite Gateway cassette. A third Multisite Gateway recombination reaction, combining the entry vectors with 8xLacO, DR5v2 and GUS, resulted in the final T-DNA construct and was transformed to *Agrobacterium tumefaciens* strain C58C1RifR (pMP90) by electroporation.

Cell culture transformation and protein accumulation studies

Arabidopsis cell cultures were transformed, maintained, upscaled and harvested as previously described (Van Leene et al., 2015). Protein extracts for protein accumulation analysis were generated by harvesting 25 mL transgenic cultures 3 days after subculturing. Harvested cell were retched for 2x 1 min at 20 Hz. Crude protein extracts were prepared in extraction buffer (Van Leene et al., 2007) by vortexing, freezing and thawing. The soluble protein fraction was obtained by a two-step centrifugation at 14,000 rpm for 15 min at 4°C. Protein concentrations were determined using the Bradford assay. 60 µg of total protein extract was loaded on a Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad), and proteins were blotted onto a PVDF membrane using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). ARF7-3xHA and TurboID-LacI protein accumulation was detected with anti-HA (12CA5, Roche). GS^{rhino}-LacI accumulation was detected with Peroxidase Anti-Peroxidase Soluble Complex antibody (Sigma). GUS accumulation was detected with anti-B-glucuronidase (Sigma).

NAA treatment and RT-PCR

Arabidopsis cell cultures transformed with the DR5v2 T-DNA construct were treated 3 days after subculturing with 100 µM PEO-IAA for one hour. PEO-IAA was removed by washing the cells with MSMO (without NAA). Cells were resuspended in MSMO containing 0,5 µg/mL NAA and grown for 4 hours. As a negative control PEO-IAA blocked cells were further incubated in MSMO, without NAA and with 100 uM PEO-IAA. Samples were taken at different time-points for RT-PCR analysis. RNA was extracted using Trizol (Invitrogen) and the RNeasy mini kit (Qiagen). cDNA was generated using the iScript cDNA synthesis kit (Bio-Rad). cDNA concentrations were measured using the Qubit dsDNA high-sensitivity assay (ThermoFisher Scientific). qRT-PCR was performed with the LightCycler 480 Real-Time SYBR Green PCR System (Roche), used primers are listed in supplementary table 1, including the housekeeping gene ACTIN 2 (AT3G18780).

Cross-linking, harvesting and DNA purification

Cross-linking of Arabidopsis cell suspension cultures was obtained using formaldehyde 3 days after subculturing. Cultures were treated for 10 min with 0,75% formaldehyde or 5 min with 0,50% formaldehyde followed by quenching the cross-linking reaction with 250 mM glycine for 10 min. Cross-linked cells were harvested as previously described (Van Leene et al., 2007). To test DNA purification tolerance under different cross-linking conditions, 1 gram of cells were ground to homogeneity in liquid nitrogen. Four volumes of ChAP-MS buffer without detergents (50 mM HEPES pH7.5, 150 mM NaCL, 0,5 mM EDTA, 1 µM trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E64), 1 mM PMSF, 0.1 mM sodium vanadate, 1 mM NaF, complete ultra EDTA free tablet (Roche)) was added followed by mixing with an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC) at 4°C. Detergents (1% Triton X-100, 0.1% NaDoc, 0.1% SDS) were added and the crude extract was incubated for 30 min at 4°C. For

decross-linking 3 volumes of water, 40 µL 5M NaCl and 0,6 µL RNase was added at 250 µL crude extract and incubated overnight at 65°C. Both cross-linked and decross-linked samples were centrifuged for 5 min at 14000 rpm, followed by 50 µg proteinase K treatment of 2 hours at 50°C. DNA was purified by the fenol/chloroform/IAA purification strategy and concentrated using the Qiaquick PCR purification Kit (Qiagen). DNA concentrations were measured using a NanoDrop Microvolume Spectrophotometer (ThermoFisher Scientific) and DNA samples were loaded on an 1,2% agarose gel.

ChIP-qPCR

Crude extracts of 5 g cross-linked and non-cross-linked cells were obtained as described above. Extracts were sonicated for 2,5 min (10 sec ON, 20 sec OFF), followed by two consecutive rounds of centrifugation at 16000 rpm at 4°C for 20 min. The extract was passed through a GF-prefilter and 0.45-µm filter (Sartorius AG), and protein concentration was determined using the Bradford assay. An input sample was kept aside, while the rest was incubated overnight with 50 µL anti-HA MagBeads (Pierce) at 4°C. Beads were washed with 10 mL of ChAP-MS buffer, 1 mL low salt buffer (20 mM Tris pH8; 150 mM NaCl; 2 mM EDTA; 1% TritonX; 0,1% SDS), 1 mL high salt buffer (20 mM Tris pH8; 500 mM NaCl; 2 mM EDTA; 1% TritonX; 0,1% SDS) and 1 mL LiCl buffer (10 mM Tris pH8; 0,25 M LiCl; 1 mM EDTA; 1% NaDoc; 1% NP-40). For decross-linking, 10 volumes Dex-buffer (10 mM Tris pH8, 1mM EDTA, 0.5M NaCl, 1% SDS) + 0.5 µl RNAse A was added, while the input sample was decross-linked as describe above. DNA purification was performed as describe above, and DNA concentrations were measured using the Qubit dsDNA high-sensitivity assay (ThermoFisher Scientific). qRT-PCR was performed with the LightCycler 480 Real-Time SYBR Green PCR System (Roche), used primers are listed in supplementary table 1, including the housekeeping gene ACTIN 2 (AT3G18780).

Coupling rabbit IgG to magnetic beads

Rabbit IgG antibodies (Sigma) were coupled in-house on BcMag™ Epoxy-activated Magnetic beads (Bioclone) as previously described (Hamperl et al., 2014).

Chromatin affinity purification (ChAP)

20 g cross-linked and/or non-cross-linked cells were ground to homogeneity in liquid nitrogen. Four volumes of ChAP-MS buffer without detergents was added followed by mixing with an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC) at 4°C. Detergents (1% Triton X-100, 0.1% NaDoc, 0.1% SDS) were added and the crude extracts were incubated on a horizontal shaker for 30 min at 4°C. Extracts were sonicated for 3 min (10 sec ON, 20 sec OFF), and supernatants were separated from cell debris by centrifugation at 2000g for 10 min at 4°C. The extracts were passed through a GF-prefilter and 0.45-µm filter (Sartorius AG), and protein concentrations were determined using the Bradford assay. The total

extracts were incubated with 300 µL in-house prepared magnetic IgG beads overnight at 4°C under gentle rotation. Beads were washed with 10 mL of ChAP-MS buffer, 2x 750 µL low salt buffer, 2x 750 µL high salt buffer and 2x 750 µL LiCl buffer. 10% of the beads were kept aside for DNA purification as described above, and purified DNA was analyzed with qRT-PCR for enrichment of the promoter of interest. The other 90% beads were subjected to 0.5 N ammonium hydroxide / 0.5 mM EDTA for 30 min at room temperature to elute proteins. The eluates were lyophilized, and the protein pellet was dissolved in 30 µL 1x NuPAGE buffer (ThermoFisher Scientific) and subjected to in-gel trypsin digestion as previously described (Van Leene et al., 2015).

For the quantitative analysis between the three promoters (DR5v2, pETG1 and pCycB1;2), on-bead digestion was performed as follows. After overnight incubation, magnetic beads were washed with 10 mL ChAP-MS buffer and 3 x 800 µL 50 mM HEPES pH7.5 + 150 mM NaCl. Beads were divided in 10% for DNA purification and 90% for on-bead digestion. On-bead digestion was preceded with three consecutive washing steps of 800 µL 50 mM ammonium bicarbonate pH8.0. Beads were resolved in 50 µl of 50 mM ammonium bicarbonate pH8.0 and 4 µl Trypsin/LysC mix (Promega) was added followed by incubated at 37°C for 3 hours with agitation. The digest was removed from the beads, followed by addition of 2 µl of Trypsin/LysC mix to the digest and incubation overnight at 37°C with agitation. The digest was centrifuged at maximum speed for 5 min and acidified to 1% TFA. Protein samples were desalted with C18 Omix tips (Agilent), which were first equilibrated with a pre-wash buffer (80% acetonitrile (AcN), 0.1% FA) and wash buffer (0.1% FA). Peptides were loaded on the C18 matrix and washed with wash buffer. Peptides were eluted in elution buffer (60% acetonitrile (AcN), 0.1% FA) and lyophilized for LC-MS/MS analysis.

Proximity labeling with TurboID

Cell cultures treated with 50 µM biotin for 24 hours at 25°C two days after subculturing were harvested as previously described (Van Leene et al., 2015). 9 g harvested cells were ground to homogeneity in liquid nitrogen and added to 6 mL of extraction buffer (100mM Tris pH7.5, 2% SDS, 8 M urea). Cells were mechanically disrupted by three repetitive freeze-thaw cycles, freezing in liquid nitrogen for 5 min and thawing in tap water for 25 min. Samples were subsequently sonicated (25 sec ON, 35 sec OFF, 25 sec ON) and incubated for 1 hour under gentle rotation at room temperature. Supernatants were separated from cell debris by two consecutive centrifugation steps at 20000 rpm for 20 min at room temperature. The extracts were passed through a GF-prefilter and 0.45-µm filter (Sartorius AG) and excess of free biotin was removed on a PD 10 Desalting Columns (Merck), which was first equilibrated with binding buffer (100 mM Tris pH7.5, 2% SDS, 7.5 M urea). Proteins were eluted from the PD 10 Desalting Columns with 3.5 mL extraction buffer. Extracts were divided in 3 experimental repeats and incubated overnight at room temperature with 100 µL Streptavidin Sepharose High Performance beads (Amersham). Supernatant was removed by centrifugation at 1500 rpm for 1 min and subsequent transfer to a mobicol column (Mo Bi Tec). Beads were washed with 4 mL of binding buffer, 800 µL of high salt

buffer (1 M NaCl, 100 mM Tris-HCl pH7.5) for 30 min and 2x 800 µL of ultrapure water. On-bead digestion was preceded with a washing step of 3.2 mL 50 mM ammonium bicarbonate pH8.0. Beads were resolved in 200 µl of 50 mM ammonium bicarbonate pH8.0 and 4 µl Trypsin/LysC mix (Promega) was added followed by incubated at 37°C overnight with agitation. An additional 2 µL Trypsin/LysC mix was added for another 2 hours at 37°C. After centrifugation at 1500 rpm for 1 min, the digest was transferred to an Eppendorf and beads were washed with 150 µL of HPLC grade water. Digest and wash sample were pooled together. Biotinylated peptides were eluted from the beads in two consecutive elution step in 300 µL solution of 0.2 % TFA, 0.1% FA and 80 % acetonitrile. Eluted peptides were lyophilized and resolved in the trypsin digest. Peptides were desalted with C18 Omix tips (Agilent) as described above and lyophilized for LC-MS/MS analysis.

Mass Spectrometry and data analysis

All ChAP-MS experiments were analyzed on a Q Exactive HF (ThermoFisher Scientific) as follows. The peptides were re-dissolved in 20 µl loading solvent A (0.1% TFA in water/ACN (98:2, v/v)) of which 10 µl was injected for LC-MS/MS analysis on an Ultimate 3000 RSLCnano system in-line connected to a Q Exactive HF mass spectrometer (Thermo). Trapping was performed at 10 µl/min for 4 min in loading solvent A on a 20 mm trapping column (made in-house, 100 µm internal diameter (I.D.), 5 µm beads, C18 Reprosil-HD, Dr. Maisch, Germany). The peptides were separated on an in-house produced column (75 µm x 400 mm), equipped with a laser pulled electrospray tip using a P-2000 Laser Based Micropipette Puller (Sutter Instruments), packed in-house with ReproSil-Pur basic 1.9 µm silica particles (Dr. Maisch). The column was kept at a constant temperature of 50°C. Peptides eluted using a non-linear gradient reaching 30% MS solvent B (0.1% FA in water/acetonitrile (2:8, v/v)) in 105 min, 55% MS solvent B in 145 min and 99% MS solvent B after 150 min at a constant flow rate of 300 nl/min. This was followed by a 10-minutes wash at 99% MS solvent B and re-equilibration with MS solvent A (0.1% FA in water). The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the 16 most abundant ion peaks per MS spectrum. Full-scan MS spectra (375-1500 m/z) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target value of 3,000,000. The 16 most intense ions above a threshold value of 13,000 were isolated (isolation window of 1.5 m/z) for fragmentation at a normalized collision energy of 28% after filling the trap at a target value of 100,000 for maximum 80 ms. MS/MS spectra (145-2,000 m/z) were acquired at a resolution of 15,000 in the Orbitrap analyzer.

For the pK8:pROLD-GS^{rhino}-LacI-35S-ARF7-3xHA-8xLacO-DR5v2-GUS samples proteins were identified with MaxQuant using parameters as listed previously (Van Leene et al., 2019). For quantitative identification of specific interactors, MaxQuant LFQ values were analyzed in Perseus as follows. First, LFQ values were Log2 transformed and potential contaminants and reverse hits were removed. Replicates were then grouped and proteins were filtered for at least two identifications in at least one group. Missing values were imputed from a normal distribution around the detection limit per sample

(width = 0.3 and down shift = 1.8). Triplicate promoter ChAP pull-downs were analyzed against a corresponding control dataset covering ChAP pull-down experiments performed on the other promoters and wild type samples. Significantly enriched proteins were identified through a volcano plot analysis with tresholds FDR = 0.01, S0 = 1.

TurboID-ChAP-MS experiments were analyzed on a Q Exactive (ThermoFisher Scientific) as previously reported (Nelissen et al., 2015). Proteins were identified with MaxQuant using parameters as listed previously (Van Leene et al., 2019). For quantitative identification of specific interactors, MaxQuant LFQ values were analyzed in Perseus as follows. First, LFQ values were Log2 transformed and potential contaminants and reverse hits were removed. Replicates were then grouped and proteins were filtered for at least two identifications in at least one group. Missing values were imputed from a normal distribution around the detection limit per sample (width = 0.3 and down shift = 1.8). Triplicate DR5v2 streptavidin pull-downs were analyzed against the corresponding control dataset covering triplicate min35S streptavidin pull-downs. Significantly enriched proteins were identified through a volcano plot analysis with parameters FDR = 0.5, S0 = 0.08.

References

Antao JM, Mason JM, Dejardin J, Kingston RE (2012) Protein landscape at Drosophila melanogaster telomere-associated sequence repeats. Mol Cell Biol **32:** 2170-2182

Bartocci C, Diedrich JK, Ouzounov I, Li J, Piunti A, Pasini D, Yates JR, 3rd, Lazzerini Denchi E (2014) Isolation of chromatin from dysfunctional telomeres reveals an important role for Ring1b in NHEJ-mediated chromosome fusions. Cell Rep **7:** 1320-1332

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Byrum SD, Raman A, Taverna SD, Tackett AJ (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep **2:** 198-205

Causier B, Ashworth M, Guo W, Davies B (2012) The TOPLESS interactome: a framework for gene repression in Arabidopsis. Plant Physiol **158:** 423-438

Chen C, Bundschuh R (2014) Quantitative models for accelerated protein dissociation from nucleosomal DNA. Nucleic Acids Res **42:** 9753-9760

Dai X, Bai Y, Zhao L, Dou X, Liu Y, Wang L, Li Y, Li W, Hui Y, Huang X, Wang Z, Qin Y (2018) H2A.Z Represses Gene Expression by Modulating Promoter Nucleosome Structure and Enhancer Histone Modifications in Arabidopsis. Mol Plant **11:** 635

Deepanksha Arora NBA, Chen Liu, Petra Van Damme, Lam Dai Vu, Anna Tornkvist, Francis Impens, Dominique Eeckhout, Alain Goossens, Geert De Jaeger, Thomas Ott, Panagiotis Moschou, Daniel Van Damme (2019) Establishment of Proximity-dependent Biotinylation Approaches in Different Plant Model Systems. BioRxiv

Dejardin J, Kingston RE (2009) Purification of proteins associated with specific genomic Loci. Cell **136:** 175-186

Fujita T, Fujii H (2011) Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS One **6:** e26109

Fujita T, Fujii H (2012) Efficient isolation of specific genomic regions by insertional chromatin immunoprecipitation (iChIP) with a second-generation tagged LexA DNA-binding domain. Advances in Bioscience and Biotechnology **3:** 626-629

Fujita T, Fujii H (2013) Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun **439:** 132-136

Fujita T, Fujii H (2013) Locus-specific biochemical epigenetics/chromatin biochemistry by insertional chromatin immunoprecipitation. ISRN Biochem **2013:** 913273

Fujita T, Kitaura F, Fujii H (2015) A critical role of the Thy28-MYH9 axis in B cell-specific expression of the Pax5 gene in chicken B cells. PLoS One **10:** e0116579

Gao XD, Rodriguez TC, Sontheimer EJ (2019) Adapting dCas9-APEX2 for subnuclear proteomic profiling. Methods Enzymol **616:** 365-383

Gao XD, Tu LC, Mir A, Rodriguez T, Ding Y, Leszyk J, Dekker J, Shaffer SA, Zhu LJ, Wolfe SA, Sontheimer EJ (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods

Guilfoyle TJ, Hagen G (2007) Auxin response factors. Curr Opin Plant Biol **10:** 453-460

Hamperl S, Brown CR, Perez-Fernandez J, Huber K, Wittner M, Babl V, Stockl U, Boeger H, Tschochner H, Milkereit P, Griesenbeck J (2014) Purification of specific chromatin domains from single-copy gene loci in Saccharomyces cerevisiae. Methods Mol Biol **1094:** 329-341

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods **3:** 11

Hayashi K, Neve J, Hirose M, Kuboki A, Shimada Y, Kepinski S, Nozaki H (2012) Rational design of an auxin antagonist of the SCF(TIR1) auxin receptor complex. ACS Chem Biol **7:** 590-598

Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem **290:** 26404-26411

Hoffmann M, Hentrich M, Pollmann S (2011) Auxin-oxylipin crosstalk: relationship of antagonists. J Integr Plant Biol **53:** 429-445

Hoshino A, Fujii H (2009) Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. J Biosci Bioeng **108:** 446-449

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

Levens D, Howley PM (1985) Novel method for identifying sequence-specific DNA-binding proteins. Mol Cell Biol **5:** 2307-2315

Lewis LK, Harlow GR, Gregg-Jolly LA, Mount DW (1994) Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in Escherichia coli. J Mol Biol **241:** 507-523

Li SB, Xie ZZ, Hu CG, Zhang JZ (2016) A Review of Auxin Response Factors (ARFs) in Plants. Front Plant Sci **7:** 47

Liao CY, Smet W, Brunoud G, Yoshida S, Vernoux T, Weijers D (2015) Reporters for sensitive and quantitative measurement of auxin response. Nat Methods **12:** 207-210, 202 p following 210

Liu X, Zhang Y, Chen Y, Li M, Shao Z, Zhang MQ, Xu J (2018) CAPTURE: In Situ Analysis of Chromatin Composition of Endogenous Genomic Loci by Biotinylated dCas9. Curr Protoc Mol Biol**:** e64

Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, Dickerson KE, Xie S, Hon GC, Xuan Z, Zhang MQ, Shao Z, Xu J (2017) In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell **170:** 1028-1043 e1019

Mair A, Xu SL, Branon TC, Ting AY, Bergmann DC (2019) Proximity labeling of protein complexes and cell type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife **8**

Matzke AJM, van der Winden J, Matzke M (2003) Tetracycline operator/repressor system to visualize fluorescence-tagged T-DNAs in interphase nuclei of Arabidopsis. Plant Molecular Biology Reporter **21:** 9-19

Mi H, Muruganujan A, Casagrande JT, Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. Nat Protoc **8:** 1551-1566

Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat Methods

Myers SA, Wright J, Zhang F, Carr SA (2017) CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus. Molecular & Cellular Proteomics **16:** S63-S63

Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, Van Lijsebettens M, Goormachtig S, Vandepoele K, Van Leene J, Muszynski M, Gevaert K, Inze D, De Jaeger G (2015) Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf. Plant Cell **27:** 1605-1619

Newell CA, Gray JC (2010) Binding of lac repressor-GFP fusion protein to lac operator sites inserted in the tobacco chloroplast genome examined by chromatin immunoprecipitation. Nucleic Acids Res **38:** e145

Pauwels L, Goossens A (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. Plant Cell **23:** 3089-3100

Qiu W, Xu Z, Zhang M, Zhang D, Fan H, Li T, Wang Q, Liu P, Zhu Z, Du D, Tan M, Wen B, Liu Y (2019) Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. Nucleic Acids Res **47:** e52

Riggs AD, Suzuki H, Bourgeois S (1970) Lac repressor-operator interaction. I. Equilibrium studies. J Mol Biol **48:** 67-83

Saha K, Bender F, Gizeli E (2003) Comparative study of IgG binding to proteins G and A: nonequilibrium kinetic and binding constant determination with the acoustic waveguide device. Anal Chem **75:** 835-842

Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H (2016) Determination of local chromatin composition by CasID. Nucleus **7:** 476-484

Sun J, Xu Y, Ye S, Jiang H, Chen Q, Liu F, Zhou W, Chen R, Li X, Tietz O, Wu X, Cohen JD, Palme K, Li C (2009) Arabidopsis ASA1 is important for jasmonate-mediated regulation of auxin biosynthesis and transport during lateral root formation. Plant Cell **21:** 1495-1511

Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inze D, De Veylder L (2008) The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. EMBO J **27:** 1840-1851

Trenner J, Poeschl Y, Grau J, Gogol-Doring A, Quint M, Delker C (2017) Auxin-induced expression divergence between Arabidopsis species may originate within the TIR1/AFB-AUX/IAA-ARF module. Journal of Experimental Botany **68:** 538-551

Van Leene J, Blomme J, Kulkarni SR, Cannoot B, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Vercruysse L, Vanden Bossche R, Heyndrickx KS, Vanneste S, Goossens A, Gevaert K, Vandepoele K, Gonzalez N, Inze D, De Jaeger G (2016) Functional characterization of the Arabidopsis transcription factor bZIP29 reveals its role in leaf and root development. J Exp Bot **67:** 5825-5840

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, Martens L, Witters E, Gevaert K, De Jaeger G (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc **10:** 169-187

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, Henderickx K, De Vijlder T, Abdelkrim A, Pharazyn A, Van Onckelen H, Inze D, Witters E, De Jaeger G (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics **6:** 1226-1238

Verkest A, Abeel T, Heyndrickx KS, Van Leene J, Lanz C, Van De Slijke E, De Winne N, Eeckhout D, Persiau G, Van Breusegem F, Inze D, Vandepoele K, De Jaeger G (2014) A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification. Plant Physiol **164:** 1122-1133

Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Wahls WP, Taverna SD, Raney KD, Tackett AJ (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics **9:** 1207-1211

Wang P, Byrum S, Fowler FC, Pal S, Tackett AJ, Tyler JK (2017) Proteomic identification of histone post-translational modifications and proteins enriched at a DNA double-strand break. Nucleic Acids Res **45:** 10923-10940

Wang R, Estelle M (2014) Diversity and specificity: auxin perception and signaling through the TIR1/AFB pathway. Curr Opin Plant Biol **21:** 51-58

Wegner GJ, Lee HJ, Corn RM (2002) Characterization and optimization of peptide arrays for the study of epitope-antibody interactions using surface plasmon resonance imaging. Anal Chem **74:** 5161-5168

Yang L, Biswas ME, Chen P (2003) Study of binding between protein A and immunoglobulin G using a surface tension probe. Biophys J **84:** 509-522

Zeng Z, Jiang J (2016) Isolation and Proteomics Analysis of Barley Centromeric Chromatin Using PICh. J Proteome Res **15:** 1875-1882

Supplementary data

Supplementary table 1: qPCR primers used for enrichment analysis of specific promoter sequences and expression level of GUS in transgenic Arabidopsis cell suspension cultures.

Name	Sequence
DR5v2 F	CCCGCGGTGTCATCTATGTG
DR5v2 R	TAGAGGAAGGGTCTTGCGGG
pCycB1;2 F	CGGCAACAGATGAAATCCCCA
pCycB1;2 R	GTCTCTGCGATTGTGACGAAG
pETG1 F	GAAACGACATCGTATGGAGAGG
pETG1 R	GAGTCTTTGCTCAAACACGAATTAAG
ACT2 F	TTGACTACGAGCAGGAGATGG
ACT2 R	ACACGAGGGCTGGAACAAG
GUS F	TACGTCCTGTAGAAACCCCAA
GUS R	CACAGTTTTCGCGATCCA

Chapter 4

CRISPR-ChAP-MS in plants:

a challenge.

Caroline Matthijs[1,2], Evi Ceulemans[1,2], Nancy De Winne[1,2], Eveline Van De Slijke[1,2],

Dominique Eeckhout[1,2], Francis Impens[3], Jelle Van Leene[1,2], Geert De Jaeger[1,2]

Ghent *University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]*

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

VIB Proteomics Core, 9000 Ghent, Belgium[3]

Author contributions

CM, JVL and GDJ designed the research; CM, EC, NDW and EVDS performed cloning and transformation experiments; CM and EC performed the research; mass spectrometry was performed by FI and mass spectrometry data was analyzed by DE; CM analyzed the data and wrote the chapter; JVL and GDJ supervised and complemented the writing.

Abstract

With the discovery of clustered regularly interspaced short palindromic repeats (CRISPR), miscellaneous molecular tools are being developed, making use of the specific recognition and binding of a genomic DNA sequence by the nuclease protein Cas9. CRISPR has been implemented in several gene-centered methods to explore the interactome of different genomic loci in mammalian cells and yeast. In this chapter we tested a similar gene-centered approach for the analysis of protein-DNA interactions at the multi-copy loci of *Arabidopsis* telomeres and ribosomal DNA. We generated a nuclease deactivated Cas9 fused to different affinity tags, but we were unable to specifically pull-down target loci. We demonstrate that plant telomeres are not the best choice for CRISPR targeting, and suggest an alternative approach implementing biotin-based proximity labeling.

Introduction

Different methods exist to identify the interacting partners of a specific genomic locus. The classical one is a Y1H screen, but alternative *in situ* methods are being developed. With the use of cross-linking and transgenes, isolation of a specific DNA region and its interactors can be achieved (Fujita and Fujii, 2013; Byrum et al., 2015). These methods, called iChIP in mammalian cells and ChAP-MS in yeast, insert LexA-binding motifs (LexA DE) near the DNA region of interest and express the transgene LexA fused to an affinity handle (LexA-tag). Due to the binding of the LexA-tag in proximity of the DNA region of interest, a specific pull-down is possible of that locus leading to the identification of protein interactors with mass spectrometry (MS). Although successful results have been obtained, the modus operandi is far from ideal. Applying genetic engineering by inserting transgenes and LexA-binding motifs could change the physiological chromatin structure and affect the native binding of proteins in the vicinity. It would be more opportune to pull-down the endogenous genomic region to get a more reliable outcome. Therefore, new approaches have been developed making use of engineered DNA-binding molecules like transcription activator-like (TAL) proteins and clustered regularly interspaced short palindromic repeats (CRISPR) (Waldrip et al., 2014; Fujii and Fujita, 2015).

CRISPR allows for the specific recognition and binding of a genomic DNA sequence by the nuclease protein Cas9 using a guide RNA (gRNA) (Jinek et al., 2012). This nuclease has two catalytic domains that are responsible for a double stranded break of the target sequence. Via targeted mutagenesis in these catalytic domains it is possible to create an inactive Cas9 nuclease that still recognize and bind the target sequence but does not contain the nuclease activity anymore (Mali et al., 2013). By combining this “dead” Cas9 (dCas9) with an affinity tag, pull-down of the target sequence and its interactors is feasible (figure 1). Two research groups already developed such tools in mammalian cells and yeast and are called engineered DNA binding molecule-mediated chromatin immunoprecipitation (enChIP) and CRISPR-ChAP-MS, (Fujita and Fujii, 2013, 2014; Waldrip et al., 2014). In short, the method starts with the expression of the dCas9 protein fused to a tag (typically a protein A derived domain binding IgG). A gRNA is co-expressed to guide the dCas9 to the target locus, chromatin is cross-linked, isolated and fragmented by sonication, and dCas9 bound chromatin fragments are enriched by pull-down with an affinity resin binding the tag. As a negative control chromatin enrichment is performed in absence of the gRNA, to correct for non-specific interactors and false positives. Chromatin fragments are reverse cross-linked by heating and eluted proteins are identified by ultrasensitive mass spectrometry (MS) like Q-exactive Orbitrap MS.

With this approach the single copy locus IRF-1 in mammalian cells could be specifically isolated together with chromatin related proteins, like histones, RNA helicases and specific IRF-1 associated proteins (Fujita and Fujii, 2013, 2014). In yeast, the GAL1 locus was targeted under transcriptionally active conditions, and label free quantitative MS analysis was performed to identify bound proteins and histone post-translational modifications (PTMs) (Waldrip et al., 2014).

In analogy with these CRISPR based pull-downs in mammalian and yeast cells, we tested if this approach could also be applicable in plant cells. We determined expression levels, localization and cross-linking capability of different tagged dCas9 proteins in our *Arabidopsis* cell suspension culture (PSB-D) before executing a pull-down assay for several multi-copy loci.

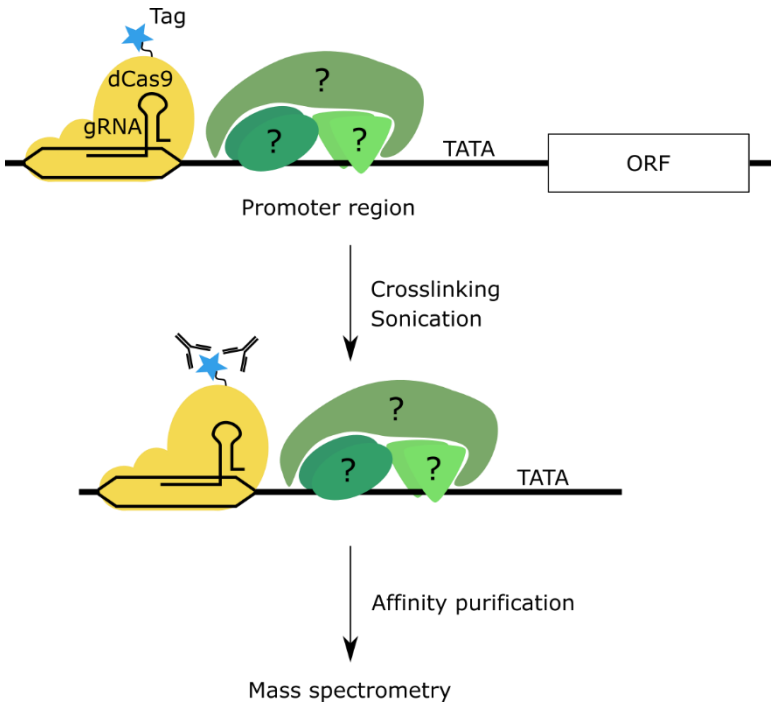


Figure 1: Scheme of CRISPR-ChAP-MS. Representation of the strategy used in enChIP (Fujita and Fujii, 2013) and CRISPR-ChAP-MS (Waldrip et al., 2014), making use of a deactivated Cas9 (dCas9), fused to an affinity tag, and a locus specific gRNA to target a specific genomic region. Molecular interactions are fixated by cross-linking and chromatin is fragmented via sonication. Subsequent affinity purification allows the isolation of a specific genomic region together with its interacting proteins. After reverse cross-linking, eluted proteins are identified by ultrasensitive mass spectrometry. In this chapter, we also applied the same strategy in *Arabidopsis* cell suspension cultures

Results

Development of a functional tagged dead Cas9 in Arabidopsis cell suspension culture

All experiments were performed with the human codon optimized Cas9 containing a C-terminal SV40 nuclear localization signal (NLS) (Mali et al., 2013). This humanized Cas9 (hCas9), successful in editing the human genome, has also be efficiently used in different plant species (Nekrasov et al., 2013; Zhou et al., 2015). Through site directed mutagenesis (D10A, H840A) the hCas9 was made catalytically inactive according to Qi et al. (2013) followed by sequencing to guarantee a successful replacement of the corresponding base pairs. The dhCas9 was fused to different affinity handles, namely the tandem affinity purification (TAP) tags GS^{rhino} (Van Leene et al., 2015) and GS^{yellow} (combination of the fluorescent protein YFP and streptavidin-binding peptide (SBP)) (Besbrugge et al., 2018), and GFP. These tags were fused N- and C-terminal via gateway cloning and cloned under control of the 35S promoter. Because C-terminal tagging could shield the NLS signal, an extra NLS-sequence was added at the N-terminus of dhCas9 (figure 2A). Transgenic cell suspension cultures (in PSB-D, grown in the dark) were made, and expression was checked for each of the dhCas9 fusion proteins. Most of these recombinant proteins accumulated at relatively low level. This could be visualized via western blotting preceded by a pull-down (figure 2B).

Because a humanized dCas9 was used, localization was evaluated in *Nicotiana benthamiana* leaves. After *Agrobacterium* mediated infiltration of GFP- or GS^{yellow}-tagged dhCas9, localization in the nucleus was observed for the GFP N- and C-terminal tagged dhCas9 and for the GS^{yellow} N-terminal tagged dhCas9 (figure 3A). In addition, a clear nucleolar localization could be visualized for the N-terminal tagged dhCas9 (figure 3B), an observation which was also reported in human MCF7 cells (Liu et al., 2017). Based on these observations and the previous publications on CRISPR-ChAP-MS (Fujita and Fujii, 2013, 2014; Waldrip et al., 2014), subsequent experiments were performed with the N-terminal tagged dCas9 constructs.

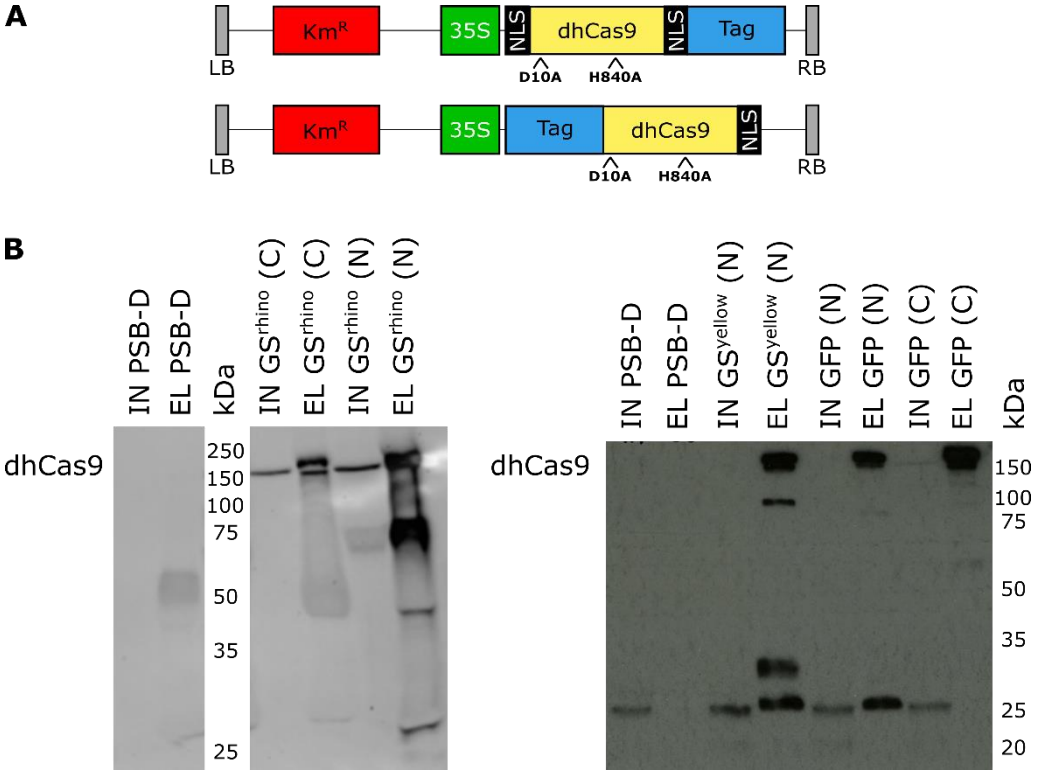


Figure 2: T-DNA constructs and expression of dhCas9 fused to GS^{rhino}, GS^{yellow} and GFP. **A)** T-DNA constructs overexpressing (35S promoter) human codon optimized dead Cas9 (dhCas9) containing a SV40 nuclear localization signal (NLS) and fused N- or C-terminal with an affinity tag. Mutations to generate a catalytically inactive Cas9 are indicated. **B)** dhCas9 C- and N-terminal tagged protein accumulation analysis. Immunoblot analysis of wild type (PSB-D) and tagged dhCas9 overexpressing cell suspension culture before (IN, 100 µg total protein) and after (EL, 25 µL total eluate) pull-down using Peroxidase Anti-Peroxidase Soluble Complex antibody (upper panel) or anti-GFP antibody (lower panel). GS^{rhino} tagged dhCas9: 180,23 kDa. GS^{yellow} tagged dhCas9: 196,63 kDa. GFP taged dhCas9: 186,6 kDa.

To be able to isolate the transient interaction between dhCas9, its target sequence and DNA-bound proteins, fixation via cross-linking is necessary. This cross-linking causes modifications to proteins and could affect the affinity for the tags. To test this, a pull-down assay was performed on cross-linked and non-cross-linked cultures containing the N-terminal tagged dhCas9 constructs. By visualizing tag-dhCas9 on western blot (figure 3C) during the different steps of the pull-down, we observed that in the unbound fraction the tagged dhCas9 could hardly be detected while detection was efficient and similar in the final eluates of both cross-linked and non-cross-linked cells, indicating that formaldehyde did not affect the affinity for the tags.

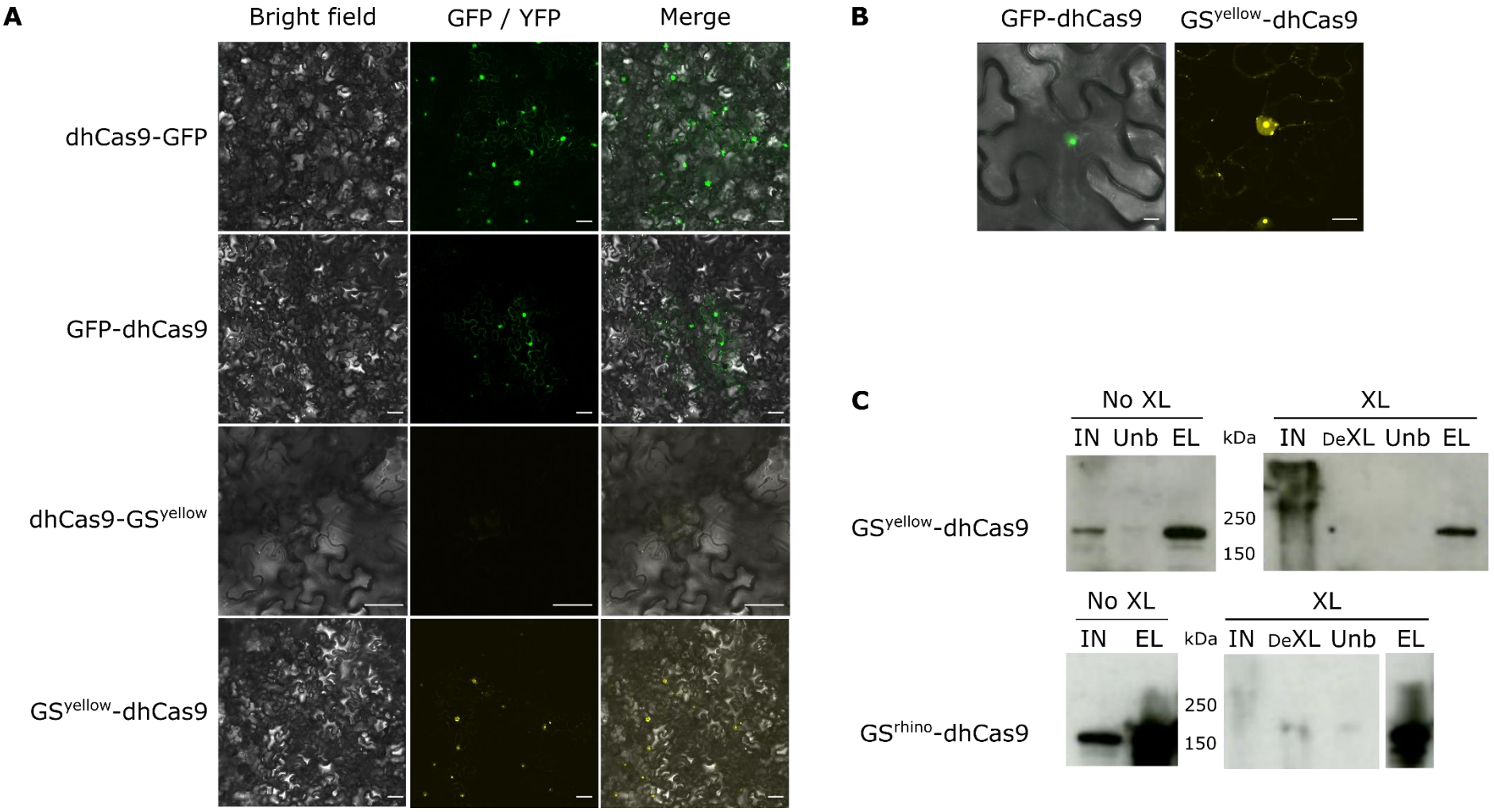


Figure 3: Nuclear localization of dhCas9 fused to GS^{rhino}, GS^{yellow} and GFP and cross-linking effect on affinity purification. **A)** Nuclear localization of tagged dhCas9 in agro infiltrated *Nicotiana benthamiana* leaf cells. Scale bar = 50 µm. **B)** Localization of tagged dhCas9 concentrated in the nucleolus in agro infiltrated *Nicotiana benthamiana* leaf cells. Scale bar = 20 µm. **C)** Cross-linking has no effect on affinity purification of tagged-dhCas9. Immunoblot analysis of GS^{yellow}- (196,63 kDa) and GS^{rhino}-dhCas9 (180,23 kDa) overexpressing cell suspension cultures during pull-down without (No XL) or with (XL) 0,75% formaldehyde cross-linking. Input samples (IN, DeXL, 60 µg total protein), unbound fraction (Unb, same volume as IN) and final eluate (EL, 1/30 of total eluate) were loaded and detection was performed using anti-GFP antibody (upper panel) or Peroxidase Anti-Peroxidase Soluble Complex antibody (lower panel).

dhCas9 interactome

Until now no AP-MS studies have been performed on the Cas9 protein. The well-known Cas9 interactions are the ones with the gRNA and the target DNA. However, there is no information about putative protein interactions. It would be informative to identify interacting proteins, not only to test if Cas9 recruits plant proteins, which could be of interest to better understand the CRISPR technique in plants, but also to identify possible non-specific binders when pulling down the DNA region of interest. The obtained list of dCas9 interactors can be used to filter out the bona fide interactors of the DNA region of interest.

TAP was performed on the GS^{rhino}-fused dhCas9 and proteins that co-eluted were analyzed by liquid chromatography-tandem MS (LC-MS/MS). Table 1 displays dhCas9 co-purified proteins after filtering out non-specific interactors and false positives (according to Van Leene et al. (2019)), among which a lot of ribosomal proteins. These proteins could reflect interactions that occur during the translation of dhCas9 mRNA and as such might not be true interactors of dhCas9, or they could reflect dhCas9 interactions happening in the nucleolus. In addition, two WD40 containing proteins were retrieved. These uncharacterized proteins contain DNA binding domains and could play a role during binding of dhCas9 on DNA. Furthermore, an insulinase was identified which is suggested to be involved in proteolysis (Rowland, 2017). When checking accumulation of GS^{rhino}-dhCas9 in our cell suspension culture, we could observe many degradation bands on western blot (figure 2B), suggesting a high degradation rate of dhCas9 in plants. This identified insulinase could be responsible for breaking down dhCas9.

Table 1: Proteins identified by LC-MS/MS after a TAP procedure from cultures producing GS^{rhino}-dhCas9 under control of the 35S cauliflower mosaic virus promoter. Proteins specifically co-purifying with the bait protein are shown only when they were confirmed in more than one experimental repeat.

Accession number	Protein name
Q99ZW2	CRISPR-associated endonuclease Cas9/Csn1
AT2G27530	60S ribosomal protein L10a-2
AT3G53740	60S ribosomal protein L36-2
AT5G02450	60S ribosomal protein L36-3
AT4G30930	50S ribosomal protein L21
AT5G23900	60S ribosomal protein L13-3
AT5G66860	Ribosomal protein L25
AT4G18905	Transducin/WD40 repeat-like superfamily protein
AT4G35370	Transducin/WD40 repeat-like superfamily protein
AT1G06900	Insulinase family protein

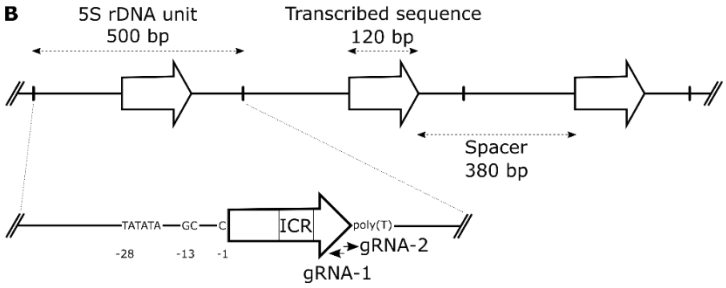
Targeting multi-copy genes

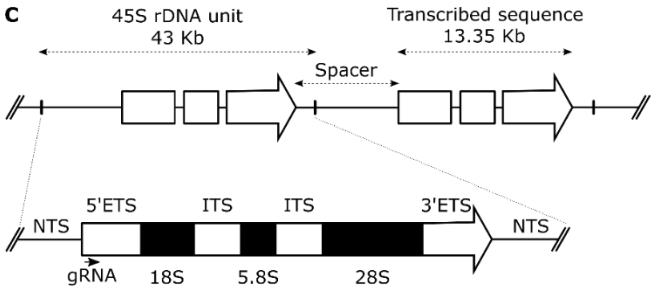
Because gene-specific protein-DNA interactions have a low abundancy and MS identification of proteins demands enough protein yield, multi-copy genes are the most interesting targets to obtain proof of concept for this CRISPR-ChAP-MS technology. Their high abundancy allows for isolation of multiple copies per cell, enriching the bona fide interactors. In *Arabidopsis,* well-known multi-copy genes are ribosomal DNA (rDNA; 45S rDNA and 5S rDNA) and the telomeres. 5S rDNA copies can be found back in the pericentromeric heterochromatin of chromosome 3, 4 and 5, while 45S rDNA loci are located on the short arms of chromosome 2 and 4 next to the telomeres. For all three genes regulatory proteins and other interacting proteins are described (Layat et al., 2012; Prochazkova Schrumpfova et al., 2019; Saez-Vasquez and Delseny, 2019). 5S rDNA transcription is executed by the RNA polymerase III, while 45S rDNA is transcribed by the RNA polymerase I. A common rDNA repressor is histone deacetylase 6 (HDA6) (Earley et al., 2010) while Target Of Rapamycin (TOR) has been shown to activate expression of 45S rDNA in *Arabidopsis* (Ren et al., 2011) and 5S rDNA in human cells and yeast (Li et al., 2006; Wei et al., 2009). Well-known telomere proteins are telomere binding protein 1 (TBP1), protection of telomeres 1a (POT1a) and telomerase (TERT) (Hwang et al., 2001; Beilstein et al., 2015)

All possible 20 nt gRNAs for all three multi-copy genes were generated using the online tool sgRNA Scorer 1.0 (Chari et al., 2015), which is a predictive model for gRNA activity based on nucleotide sequence and epigenetic parameters uncovered during a high-throughput analysis in mammalian cells. A list of possible off-targets for every gRNA was generated by Cas-OFFinder (Bae et al., 2014), an online tool which does not limit the number of mismatches and allows variations in protospacer-adjacent motif (PAM) sequences recognized by Cas9. Furthermore, extra criteria were imposed as was recommended by Liang et al. (2016), including GC content and 3D-structure of the gRNA. We also considered the position of different gRNAs relative to the binding sites of known regulatory proteins (Layat et al., 2012), resulting in the selection of gRNAs close to the internal promoter of 5S rDNA and in the 5’ external transcribed spacer (ETS) region of the 45S rDNA. Figure 4 displays the final gRNAs, their position in the target locus and possible off-targets. T-DNA constructs were generated via gateway cloning expressing the tagged dhCas9 under control of the 35S promoter and the gRNA under control of the U6 promoter. Constructs were transformed in our *Arabidopsis* cell suspension culture.

A

Target locus	gRNA sequence	Name
Telomere	GGGTTTAGGGTTTAGGGTTT	Tel-gRNA
5S rDNA	CAACACGAGGACTTCCCGGG	5S-gRNA-1
	TAAAAGAGGGATGCAACACG	5S-gRNA-2
45S rDNA	GAGGGAGTCTGGGCAGTCCG	45S-gRNA





D

DNA target 5S-gRNA-1	CAACACGAGGACTTCCCGGGAGG
Off-target Ch 5	CAACACGAGGACTTtCCGGGAGG
Off-target Ch 2	CAACACGAGGACTTCtcaGGAGG
DNA target 5S-gRNA-2	TAAAAGAGGGATGCAACACGAGG
Off-target Ch 2 (2)	aAAAAaAGGGATGCAACACGAGG
Off-target Ch 2 (4)	ccAAAaAGGGATtCAACACGAGG
DNA target 45S-gRNA	GAGGGAGTCTGGGCAGTCCGTGG
Off-target Ch 1	GAGGGAGTCTGaGCAGTCCGAGG
Off-target Ch 5	GgtGGGATTCTGGGCgGTCtGAGG

Figure 4: gRNA sequence and position for the target loci telomeres, 5S rDNA and 45S rDNA. **A)** Table with the different gRNA sequences. **B)** Position of 5S-gRNA-1 and 5S-gRNA-2 in the 3'terminal end of the transcribed region of 5S rDNA. ICR: internal control region, internal promoter. **C)** Position of 45S-gRNA in the 5' external transcribed spacer (ETS) of 45S rDNA. NTS: non-transcribed spacer. ITS: internal transcribed spacer. **D)** gRNA target sequences including the PAM (blue) and the corresponding off-targets with mismatches shown in red. Off-targets were determined via Cas-OFFinder (Bae et al., 2014).

CRISPR-ChAP-MS on telomeres

CRISPR-ChAP-MS on telomeres was performed with the GS^{rhino} and GS^{yellow} tags. A detailed overview of the pull-down protocol can be found under Material and Methods. In brief, after cross-linking, cells were lysed and chromatin was fragmented in pieces of around 500 bp via sonication (figure 6A). A suitable affinity resin (IgG beads for GS^{rhino} and anti-GFP beads for GS^{yellow}) was added to the lysate to allow capturing of the dhCas9-targeted chromatin complex. After several washing steps the complex was eluted and cross-linking was reversed. Pull-down efficiency was checked on western blot and via a PCR specifically designed for amplification of telomere sequences (Cawthon, 2009; Vaquero-Sedas and Vega-Palas, 2014).

Amplification of telomere sequences is a hurdle because of its repetitive nature, resulting in complementary primers and the formation of primer dimers. Therefore,

successful amplification of *Arabidopsis* telomere sequences has been achieved by using degenerative primers (Vaquero-Sedas and Vega-Palas, 2014). These primers are not perfectly complementary with the telomere sequence but creates sufficient hydrogen bonds for binding and amplification. Primer dimers are also formed but have a lower amount of hydrogen bonds and the 3'ends are not compatible, preventing replication by the polymerase (figure 5A). For telomere amplification a standard PCR protocol was adapted based on the monochrome multiplex quantitative PCR method (MMQPCR) (Cawthon, 2009). Two cycles of annealing at lower temperature (49°C) were included to allow binding of the degenerative primers with the telomere sequence (figure 5B). Although primer dimer amplification should be prevented by the structure of the primers, some polymerases have shown the opposite (Jodczyk et al., 2015). To avoid additional amplification of primer dimers in the presence of telomere sequences, the HOT FIREPol® (Solis Biodyne) was tested, as it has been shown to avoid primer dimer amplification during MMQPCR (Jodczyk et al., 2015). However, in our hands primer dimers are still amplified in absence of genomic DNA using HOT FIREPol® (figure 5C). In the presence of genomic DNA, telomere amplification is observed by a smear on the DNA agarose gel (figure 5C). This smear is the result of the degenerative primers binding at different positions on the telomeres resulting in the amplification of different lengths of sequences (Vaquero-Sedas and Vega-Palas, 2014). In addition, the use of HOT FIREPol® results in more efficient amplification of telomere sequences then the standard Taq polymerase (figure 5C).

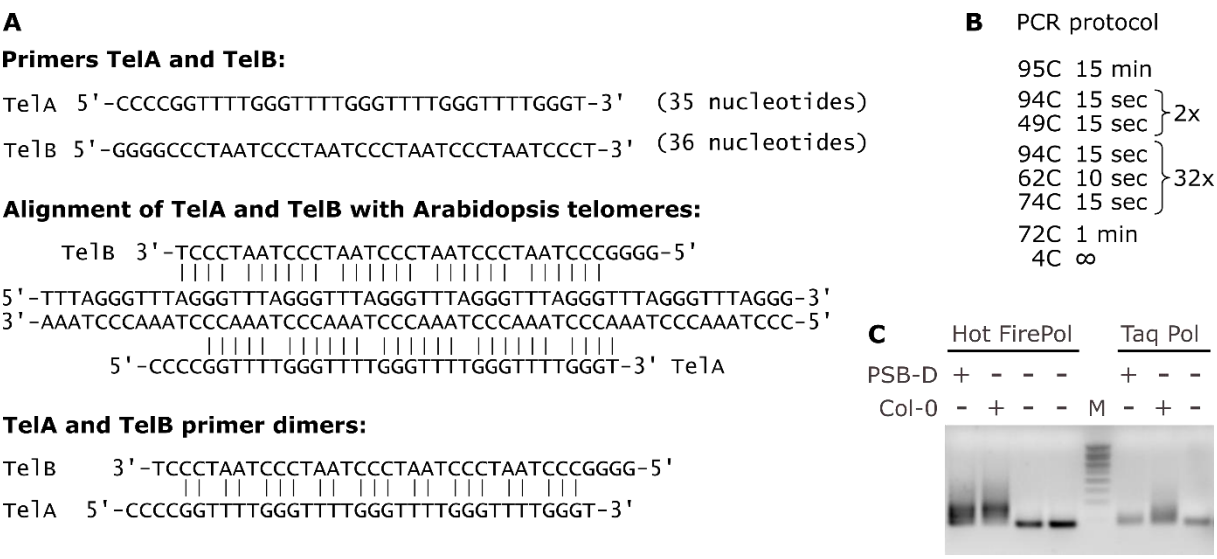


Figure 5: Telomere specific PCR. **A)** Representation of the degenerative telomere primers TelA and TelB and their complementation with the telomere sequence and primer dimer forming. Figure adopted from Vaquero-Sedas and Vega-Palas, 2014. **B)** Adopted PCR protocol for amplification of telomere sequences. **C)** PCR result for amplification of telomere sequences with genomic DNA of PSB-D and Col-0 as template. PCR was performed with two different polymerases, HOT FIREPol® (Solis Biodyne) and standard Taq polymerase. Amplification of telomeres is visualized by a smear of different telomere sequence lengths. Without template, primer dimers are formed. M: SmartLadder SF (Eurogentec).

After pull-down of the telomeres, a huge amount of dhCas9 degradation products was observed on western blot in the eluate, while this was less the case in absence of the Tel-gRNA (figure 6B). Also in the input material breakdown products are more present when tagged dhCas9 is recruited to the telomeres. Telomere sequences could be amplified by PCR in the eluate after GS^{rhino} pull-down, but this was also the case without the presence of the Tel-gRNA (figure 6C). On the other hand, telomere amplification was not observed after GS^{yellow} pull-down (figure 6D) suggesting an inefficient isolation of telomeres. In conclusion, CRISPR-ChAP on the telomeres was unsuccessful, so no MS analysis was performed.

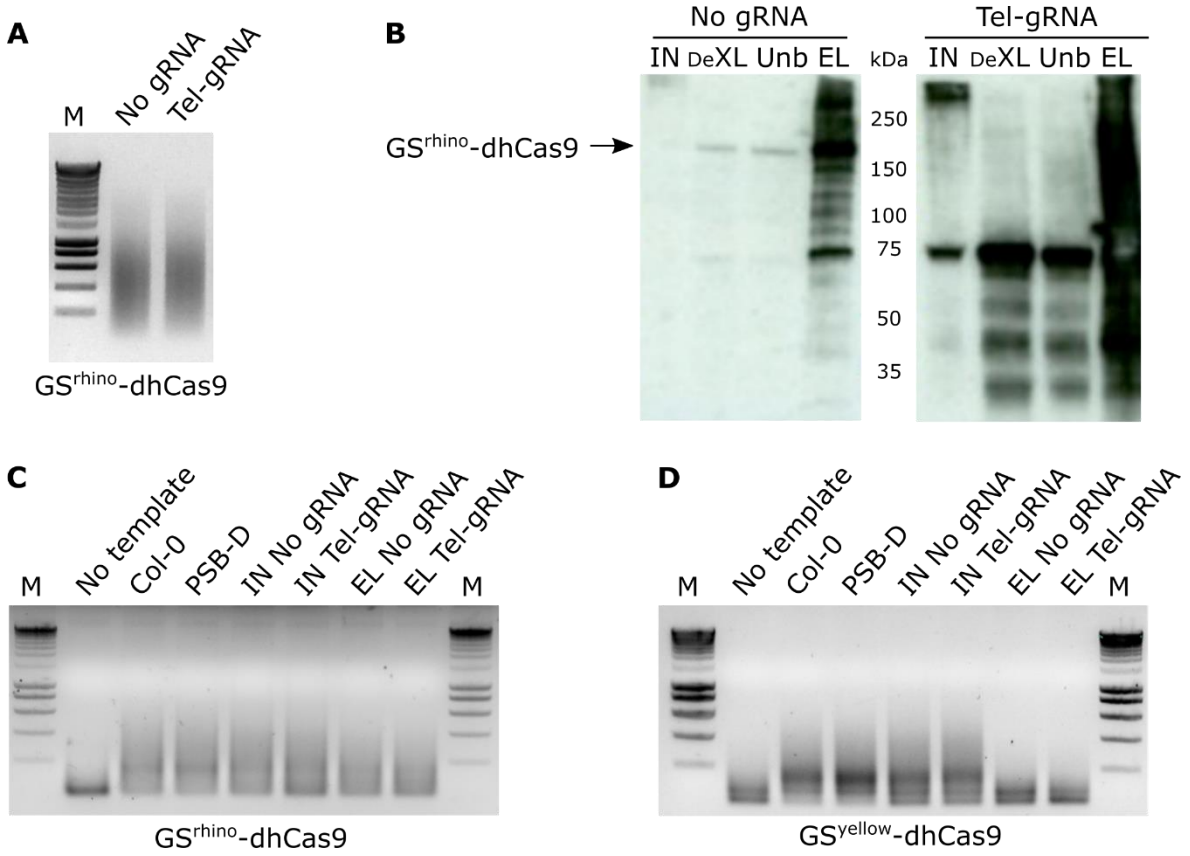


Figure 6: CRISPR-ChAP-MS on telomeres. **A)** DNA agarose gel with sonicated total DNA extract of GS^{rhino}-dhCas9 cell cultures, with or without (No gRNA) Tel-gRNA. **B)** Immunoblot analysis of GS^{rhino}-dhCas9 (180,23 kDa) during pull-down with or without (No gRNA) Tel-gRNA. Input samples (IN, DeXL, 60 µg total protein), unbound fraction (Unb, same volume as IN) and final eluate (EL, 1/60 of total eluate) were loaded and detection was performed using Peroxidase Anti-Peroxidase Soluble Complex antibody. **C-D)** PCR amplification of telomere sequences before (IN) and after (EL) pull-down with GS^{rhino}-dhCas9 (C) and GS^{yellow}-dhCas9 (D), with or without (No gRNA) Tel-gRNA. As negative control PCR was performed without template DNA. As positive control PCR was performed on genomic DNA of Col-0 and PSB-D. M: SmartLadder SF (Eurogentec).

CRISPR-ChAP-MS on rDNA

CRISPR-ChAP-MS on rDNA was performed with the GS^{rhino} tag. As mentioned before two different gRNAs were developed for targeting 5S rDNA and one gRNA for 45S rDNA (figure 4) based on different selection criteria. In a first attempt, 5S-gRNA1 was tested for isolation of 5S rDNA.

One 5S rDNA unit is 500 bp long, containing the transcribed region of 120 bp flanked by a spacer sequence of 380 bp (figure 4B). To have an efficient isolation of the 5S rDNA loci and lowering the chance of pulling down neighboring sequences, sonication conditions were established in such a way that chromatin fragments of around 500 bp were obtained (figure 7A-B). An efficient affinity purification was observed since the full length GS^{rhino}-dhCas9 protein was present in the final eluate (figure 7D). With qPCR specific isolation of the 5S rDNA locus was checked by following the amplification rate of 5S rDNA and two possible off-targets determined via the online tool Cas-OFFinder (figure 4D). Off-target one is located on chromosome 5 (position 22841072) and has one mismatch with 5S-gRNA-1. The other off-target is located on chromosome 2 (position 3562153), a chromosome without 5S rDNA, and has two mismatches with 5S-gRNA-1. Although a small enrichment of 5S rDNA was observed with two different primer sets, a 2-fold higher enrichment of the chromosome 5 located off-target was detected (figure 7G), suggesting a non-specific binding of the 5S-gRNA-1. CRISPR-ChAP with 5S-gRNA-2 resulted also in an efficient affinity purification of the GS^{rhino}-dhCas9 (figure 7E). Also here, two off-targets were analyzed for specificity of the gRNA (figure 4D), both located on chromosome 2 (position 3562140 and 10066080) containing two or four mismatches with 5S-gRNA-2. Enrichment of the off-targets was not observed, but also 5S rDNA enrichment could not be detected with qPCR (figure 7H).

45S rDNA is 43 Kb long, containing a transcribed region of 13.35 Kb (figure 4C). To identify the transcription regulatory factors, a gRNA was designed in close proximity of the promoter region, in the beginning of the 5' external transcribed spacer (ETS) and close to the TOR binding site (Ren et al., 2011). In this way, chromatin fragments of 500 bp could be generated with sonication (figure 7C), allowing the isolation of the transcription regulatory region of 45S rDNA. Analysis of two off-targets were included located on chromosome 1 (position 12293905) and 5 (position 12792053) having one and five mismatches respectively with the 45S-gRNA (figure 3D). Also here an efficient affinity purification of GS^{rhino}-dhCas9 was observed (figure 7F) but significant enrichment of the 45S rDNA regulatory region was not obtained (figure 7I). Due to these negative outcomes, MS analysis for the rDNA targets was not performed.

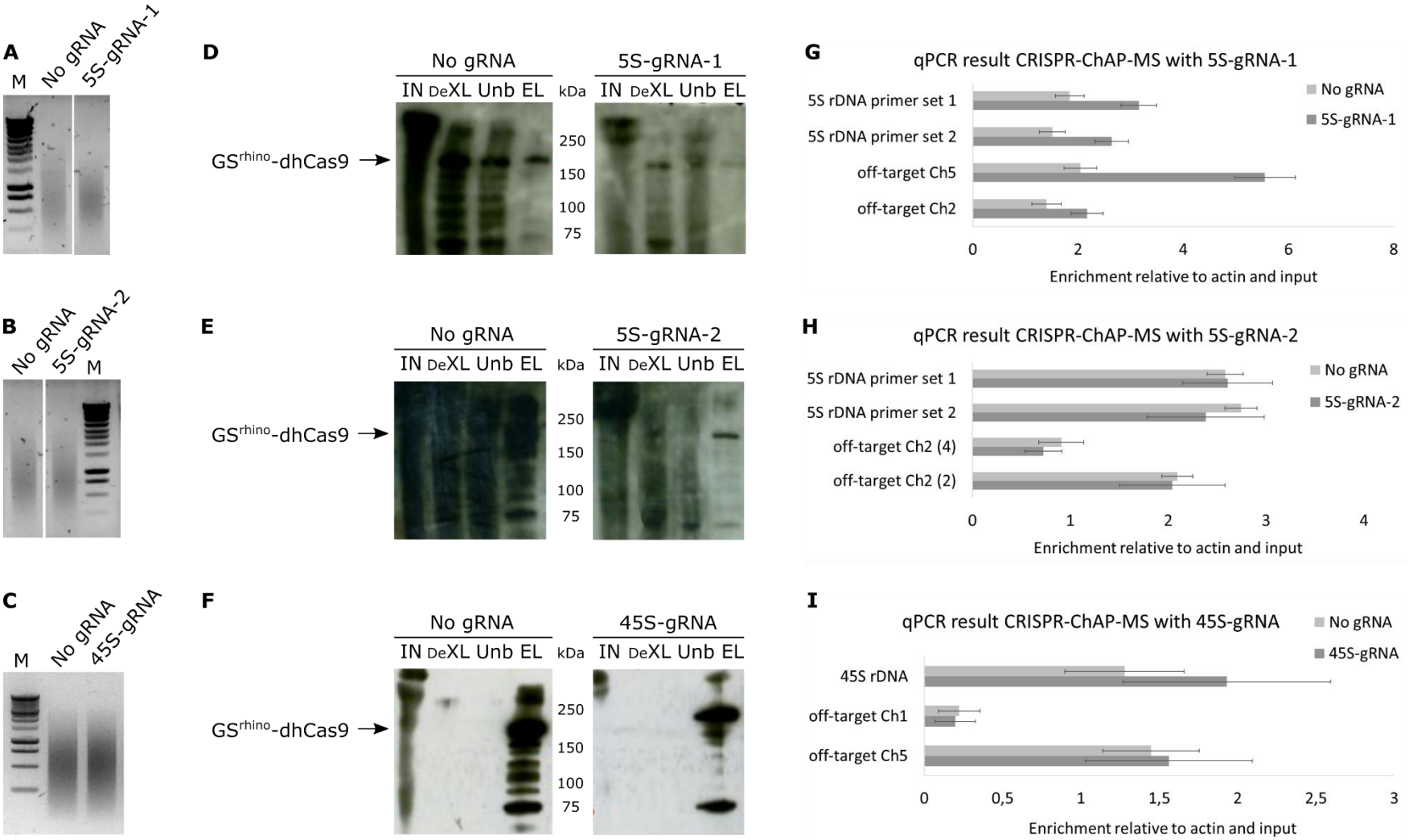


Figure 7: CRISPR-ChAP-MS on rDNA. **A-B-C)** DNA agarose gel with sonicated total DNA extract of GS^{rhino}-dhCas9 cell cultures, with or without (No gRNA) 5S-gRNA-1 (A), 5S-gRNA-2 (B) or 45S-gRNA (C). **D-E-F)** Immunoblot analysis of GS^{rhino}-dhCas9 (180,23 kDa) during pull-down with or without (No gRNA) 5S-gRNA-1 (D), 5S-gRNA-2 (E) or 45S-gRNA (F). Input samples (IN, DeXL, 60 µg total protein), unbound fraction (Unb, same volume as IN) and final eluate (EL, 1/60 of total eluate) were loaded and detection was performed using Peroxidase Anti-Peroxidase Soluble Complex antibody. **G-H-I)** qPCR analysis after pull-down with or without (No gRNA) 5S-gRNA-1 (G), 5S-gRNA-2 (H) or 45S-gRNA (I). For amplification of the 5S rDNA locus two different primer sets were used. For every gRNA two possible off-targets were analyzed on different chromosomes (Ch). Values were normalized relative to the reference gene actin and input DNA.

Discussion

Identifying the interacting proteins of a specific genomic locus can be accomplished in different way's. The classical one is a Y1H screen, but alternative *in situ* methods are on the rise like iChIP (Hoshino and Fujii, 2009; Fujita and Fujii, 2013), ChAP-MS (Byrum et al., 2012; Byrum et al., 2015) and methods implementing engineered DNA-binding molecules like TAL and CRISPR (Waldrip et al., 2014; Fujii and Fujita, 2015). Here we tested if a similar approach with CRISPR would work in plant cells. An inactive version of a human codon optimized Cas9, fused to different affinity tags was generated. Expression of tagged-dhCas9 was observed in our PSB-D cell culture, as was the localization in the nucleus with a high concentration in the nucleolus. These preliminary results suggested a functional dhCas9 in plant cells, which should have allowed us to isolate specific genomic loci when combined with a specific gRNA. However, targeting telomeres and rDNA wasn't successful, showing that further optimization of the CRISPR-ChAP-MS technique in plants is required.

A plant codon optimized Cas9

Although the human codon optimized Cas9 was already used in different plant species like *Arabidopsis*, tobacco, poplar and rice (Feng et al., 2013; Mao et al., 2013; Nekrasov et al., 2013; Xie and Yang, 2013; Zhou et al., 2015), usage of the inactive version has not yet been reported. Here we showed that also in *Arabidopsis* dhCas9 is expressed, but expression is rather low causing the need for a pull-down to visualize it on western blot. This low expression level in combination with the high degradation rate in *Arabidopsis* cells, reduces the chance of tagged-dhCas9 to bind the target locus and could lead to an inefficient pull-down, as we have observed. Increasing the expression by using a plant codon optimized Cas9 or a stronger constitutive plant promoter could be a solution. In 2014 a Cas9 was codon-optimized for *Arabidopsis thaliana* by the group of Holger Puchta and expression was driven by the constitutive Ubiquitin4–2 promoter from *Petroselinum crispum* (PcUbi4-2) (Fauser et al., 2014). Efficient non-homologous end-joining (NHEJ)-mediated mutagenesis was obtained as well as the efficient generation of heritable mutations in *Arabidopsis* plants by Agro-bacterium-mediated transformation. Adopting this Cas9, codon-optimized for *A. thaliana*, in our CRISPR-ChAP-MS approach could result in higher expression levels. An initial test demonstrates that this is indeed the case, even under the 35S promoter (supplementary figure 1). Subsequent CRISPR-ChAP-MS experiments using this dCas9 could provide the evidence for a more efficient pull-down due to the higher expression.

Stimulation of rDNA transcription

Despite the low expression, CRISPR-ChAP-MS experiments were performed on multi-copy targets. Multi-copy genes were chosen because of the limited detection sensitivity of the MS. Specific protein-DNA interactions are not that abundant in a cell, causing the need for a high amount of starting material to obtain enough protein yield for detection with MS (which is at least a picomole of protein). To circumvent this, one can target multi-copy genes where specific protein-DNA interactions are repeatedly present, increasing the protein yield after pull-down. Telomere sequences and rDNA are highly abundant in the *Arabidopsis* genome, and have been studied extensively resulting in the identification of multiple regulatory and interacting proteins. Furthermore, transcription of 45S rDNA takes place within the nucleolus, the cellular place where tagged-dhCas9 is most abundant. Additionally, our *Arabidopsis* cell suspension culture has a ploidy level of 9C (unpublished data), which increases the number of a pulled-down specific target sequence. This makes multi-copy loci promising targets for testing out the CRISPR-ChAP-MS technique in our cell suspension culture. However, pull-down of these specific genomic regions could not be obtained. One main reason could be the low expression level of the tagged-dhCas9, as mentioned before. Another bottleneck could be the native chromatin structure.

Hundreds of ribosomal RNA (rRNA) genes are present in the *Arabidopsis* genome but only a subset is active. rDNA is subjected to a dosage control that regulates the number of active rRNA genes depending on the cellular demands for ribosomes and protein synthesis (Lawrence and Pikaard, 2004). Epigenetic regulation of rDNA, like histone methylation, ensures that most of these rRNA genes are situated in a heterochromatin state. This condensed chromatin however is not receptive to the binding of proteins or complementation with RNA molecules, which is essential in our CRISPR-ChAP-MS protocol. This selectivity towards an active chromatin state was also observed when enChIP and CRISPR-ChAP-MS were applied in mammalian cells and yeast (Fujita and Fujii, 2014; Waldrip et al., 2014). Both needed to activate transcription of their target gene to be able to do a successful pull-down. Fujita and Fujii induced transcription of the IRF-1 by treating the mammalian cells with interferon (IFN) γ, while Waldrip et al. grew yeast on galactose to activate the GAL1 locus. In order to have a successful pull-down of rDNA, we must attempt to activate more rRNA gene transcription by increasing the cellular demands for ribosomes and protein synthesis. One way could be by depleting our PSB-D cell culture of carbohydrates, resulting in an arrest of cell growth, after which carbohydrates are added again, leading to the reactivation of cell growth and a boost in protein synthesis. Another way is by adding a stimulus to the cell culture. It has been postulated that the plant hormone cytokinin could also stimulate rRNA gene

expression, as it has been shown to up-regulate transcription initiation from the polymerase I promoter in *Arabidopsis* (Gaudino and Pikaard, 1997; Koukalova et al., 2005). Extra addition of cytokinin to the PSB-D culture, could be a solution.

CRISPR-unfriendly telomeres

Telomeres are non-transcribed sequences at the end of the chromosomes, protecting them from deterioration or from fusion with neighboring chromosomes. They are bound by different proteins, some of which guard chromosome ends and regulate telomerase access (Nelson and Shippen, 2012). Fujita et al. were able to isolate telomere sequences in mammalian cells, using their engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) method (Fujita et al., 2013). They identified known and novel telomere-binding proteins as well as known telomere-binding RNAs, including the telomerase RNA component (Terc), and a number of novel telomere-binding non-coding RNAs (Fujita et al., 2015). To achieve this, they used a transcription activator-like (TAL) protein recognizing telomere repeats. In 2016 CRISPR was used for targeting and investigating the local telomere chromatin environment in mammalian cells, a technique called CasID (Schmidtmann et al., 2016). Here a dCas9 was fused with the promiscuous biotin ligase BirA* and in combination with a telomere specific gRNA, proteins in the direct vicinity were biotinylated. Subsequent streptavidin-mediated precipitation and mass spectrometry identified components of the telomere specific shelterin complex among others. One year later, an alternative purification strategy with CRISPR was developed making use of a biotinylated dCas9 and was called CAPTURE (CRISPR affinity purification in situ of regulatory elements) (Liu et al., 2017). Also here successful results were obtained targeting the telomere region in mammalian cells.

Because of their success, targeting plant telomeres with our CRISPR-ChAP-MS technique seemed to be a good choice. However, our results show otherwise. Specific isolation of telomere sequences could not be observed, but even more noticeable was the strong degradation rate of tagged-dhCas9 when targeted to the telomeres. It seems that the plant telomere region and its protecting role against the degradation of the 5’ chromosome ends by nucleases also has an effect on the endonuclease Cas9.

How come this telomere specific dCas9 degradation was not observed in mammalian cells during CasID and CAPTURE experiments? There is a presumption that susceptibility of telomeres to nucleases differs between organisms (Lydall, 2003). This is based on the fact that telomerase-deficient human cells lose approximately 150 bp of telomeric DNA per generation (Huffman et al., 2000), while in yeast only 3-6 bp are

lost (Lundblad and Blackburn, 1993). Meaning that human telomeres are probably more susceptible for nuclease then yeast. It has been reported that in telomerase-deficient *Arabidopsis* plants the rate of telomere loss is 10 times less than the rate reported for telomerase-deficient mice (Fitzgerald et al., 1999). This suggests that *Arabidopsis* telomeres are less susceptible for nucleases, and forces might be present inhibiting the function of nucleases including dCas9. By replacing the CRISPR components with a plant telomere-specific nuclease-free TAL protein, as Fujita et al. (2013) has done, this problem could be bypassed and specific isolation of plant telomeres could be achieved with our pull-down protocol.

Bypassing dCas9 pull-down with proximity labeling

We demonstrated that pull-down of multi-copy target loci with tagged dCas9 is a hurdle in *Arabidopsis* cell suspension culture. However, absence of target sequences after pull-down does not mean that the tagged dCas9 is not binding on the target loci. Binding could occur at such a low rate that it is undetectable after pull-down. To still allow detection of locus-specific interacting proteins, we could replace the affinity tag by an enzyme that is capable of covalently labeling proteins in the immediate vicinity. This strategy has been applied in mammalian cells and yeast using a promiscuous biotin ligase BirA (CasID) (Schmidtmann et al., 2016) or an ascorbate peroxidase (C-BERST, GloPro) (Myers et al., 2017; Gao et al., 2018) fused to dCas9. Upon biotin treatment, all proteins in the neighborhood of the target locus will be biotinylated and can be subsequently retrieved via a stringent streptavidin affinity purification. In this way, biotin-labeled proteins are pulled-down instead of a specific protein-DNA macromolecular complex. Even when the dCas9 DNA-binding event would be short, extension of the biotin treatment time-span will result in the accumulation of biotin-labeled locus specific proteins. Together with the incorporation of appropriate negative controls (e.g. without gRNA, without biotin addition,…), the efficiency of this strategy has been reported in different studies, exploring multi-copy loci and single loci (Schmidtmann et al., 2016; Myers et al., 2018; Gao et al., 2019). With the development of a new promiscuous variant of BirA, TurboID (Branon et al., 2018), which has an increased catalytic efficiency and is active under standard plant growth conditions (Mair et al., 2019), a similar approach in plants is on the horizon.

Conclusion

With the discovery of CRISPR, a new and rapidly developing approach for gene centered analysis is rising and will likely be the method of choice in the future for the

identification of protein-DNA interaction. Native loci can be targeted without changing the physiological chromatin structure. Successful gene centered experiments have been performed with CRISPR (Fujita and Fujii, 2013; Waldrip et al., 2014) and new variants are being developed which are making use of the high affinity between biotin and streptavidin (Schmidtmann et al., 2016; Liu et al., 2017; Liu et al., 2018). We tried to develop a similar gene centered technique for the identification of protein-DNA interactions in plant cells by targeting a specific native chromatin region via CRISPR. However, our initial approaches were ineffective. Further optimization is needed, by improving the stability of dCas9 in our PSB-D cell culture and by implementing a proximity-based labeling strategy.

Materials and methods

Cloning

Generation of expression constructs was obtained by following a gene stacking approach with the use of the MultiSite Gateway technology (Karimi et al., 2002). Supplementary table 1 displays an overview of the used primers. A gateway entry vector containing hCas9 (Mali et al., 2013) flanked by AttL1/AttL2 recombination sites was kindly provided by M. Karimi. A deactivated dhCas9 entry vector was obtained by two consecutive site directed mutagenesis reactions using mutated primers, the Pfu polymerase (Promega) and DpnI (New England BioLabs) to degrade the original template vector. Correct mutagenesis was checked via sequencing with several internal primers. A dhCas9 without stop codon and with N-terminal NLS sequence was amplified from the dhCas9 entry vector using the Kappa polymerase (Kapa Biosystems) flanked by AttB1/AttB2 recombination sites and subjected to a BP reaction with pDONR221 to generate the corresponding entry vector. Generation of the GS^{rhino} and GS^{yellow} entry vectors is described elsewhere (Van Leene et al., 2015; Besbrugge et al., 2018). The p35S and GFP entry vectors as well as the pK8m43GW2 and pK7m24GW2 destination vectors can be retrieved at https://gateway.psb.ugent.be/. Subsequent MultiSite Gateway recombination in the pK8m43GW2 and pK7m24GW2 destination vectors (Karimi et al., 2002) generated the expression vectors pK7:35S-GS^{yellow}-dhCas9, pK7:35S-GS^{rhino}-dhCas9, pK7:35S-GFP-dhCas9, pK8:35S-NLS-dhCas9-GS^{yellow}, pK8:35S-NLS-dhCas9-GS^{rhino}, pK8:35S-NLS-dhCas9-GFP. To insert a second Multisite Gateway cassette, the expression vectors were linearized via AvrII/XbaI restriction followed by ligation of the XbaI restricted fragment from pXb2M43GW7. Oligonucleotides (see supplementary table 1) for the different target loci were annealed to each other by 10 min incubation in a thermoblock and subsequent cooling down to room temperature. The pEN-L4-U6-gRNA-L3 vector was linearized by BsaI restriction, cutting out 13 nt between U6 promoter and gRNA backbone. Annealed oligonucleotides were ligated in the linearized pEN-L4-U6-gRNA-L3 vector, based on overlapping sequence ends. The resulted gRNA entry vectors were recombined with the dhCas9-m34GW expression vectors to generate the final T-DNA constructs. These were transformed to *Agrobacterium tumefaciens* strain C58C1RifR (pMP90) by electroporation.

Cell culture transformation and protein accumulation studies

Arabidopsis cell cultures were transformed, maintained, upscaled and harvested as previously described (Van Leene et al., 2015). Protein extracts for protein accumulation analysis were generated by harvesting 25 mL transgenic cultures 3 days after subculturing. Harvested cell were retched for 2x 1 min at 20 Hz. Crude protein extracts were prepared in extraction buffer (Van Leene et al., 2007) by vortexing, freezing and thawing. The soluble protein fraction was obtained by a two-step centrifugation at 14,000 rpm for 15 min at 4°C. Protein concentrations were determined using the Bradford assay. For

pull-down, 2 mg protein extract was incubated with 25 µL IgG-sepharose beads (for GS^{rhino}, GE Healthcare) or GFP-trap Agarose beads (for GS^{yellow} and GFP, Chromotek) overnight at 4°C. Beads were washed with 1,5 mL extraction buffer, and pulled-down proteins were eluted in 25 µL sample buffer. 100 ug op total protein extract or total volume of eluates were loaded on a Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad), and proteins were blotted onto a PVDF membrane using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). GS^{rhino}-tagged dCas9 accumulation was detected with Peroxidase Anti-Peroxidase Soluble Complex antibody (Sigma), while GS^{yellow} and GFP fusions were detected with polyclonal anti-GFP (AB290, Abcam).

Transient expression in *Nicotiana benthamiana* leaves

Transgenic *Agrobacterium tumefaciens* strains were grown in yeast extract beef (YEB) broth with 10mM MES and 20uM acetosyringone until an OD_{600} of 3. In parallel an *Agrobacterium tumefaciens* strain containing a vector encoding the silencing suppressor P19 was cultivated. Bacterial samples were centrifuged for 10 min at 4000 rpm, and the pellet was resolved in infiltration buffer (100uM acetosyringone, 10mM $MgCl_2$, 10mM MES, pH5.7). Equal amounts of dCas9 transformed and P19 containing agro's were mixed and gently infiltrated in *Nicotiana benthamiana* leaves with a syringe. Tobacco plants were grown for 3-4 days, and infiltrated areas were cut out and analyzed by confocal microscopy. Image acquisition was obtained with a 100M inverted confocal microscope (Zeiss) with software package LSM510 version 3.2 equipped with a 10x, 25x or 40x water-corrected objective using the following settings for GFP and YFP detection: GFP excitation at 488 nm; emission filter 505-530 nm, YFP excitation at 488 nm; emission filter 505-550 nm.

Tandem affinity purification and MS analysis

Tandem affinity purification of GS^{rhino}-dhCas9 was performed as previously described (Van Leene et al., 2015) with adaptation to the elution step and subsequent in-gel trypsin digest. Purified proteins were digested on-bead as follows. After the final wash with extraction buffer without detergent, the beads were washed with 500 µL 50 mM NH4HCO3 (pH 8.0). The wash buffer was removed and 50 µL 50 mM NH4OH was added together with 1 µg Trypsin/Lys-C and incubated at 37°C for 4 h in a thermomixer at 800 rpm. Next, the digest was separated from the beads and overnight incubated with 0.5 µg Trypsin/Lys-C at 37°C. Finally, the digest was centrifuged at 20800 rcf in an Eppendorf centrifuge for 5 min, the supernatant was transferred to a new 1.5 mL Eppendorf tube, and the digest was dried in a Speedvac and stored at -20°C until MS analysis. Protein samples were analyzed on a Q Exactive (ThermoFisher Scientific) as previously reported (Nelissen et al., 2015). After MS-based identification of co-purified proteins, specific proteins in the TAP experiments were detected by comparison against an updated list of non-specific proteins (Besbrugge et al., 2018), determined as previously described (Van Leene et al., 2015). True interactors that might have been missed because of their presence in the list

of non-specific proteins were retained through a semi-quantitative analysis. In this approach, normalized spectral abundance factors (NSAF) of all *Arabidopsis* proteins detected in the GS^{rhino}-dhCas9 sample were compared against the corresponding average NSAF deduced from a control TAP dataset. This control dataset was built from more than 1000 TAP experiments, covering a huge set of bait proteins not related to Cas9. For stringent filtering of specific proteins, only proteins identified with at least two peptides were retained that were highly (at least 10-fold) and significantly [-log10(p-value(T-test)) ≥10] enriched compared to the control datasets.

Coupling rabbit IgG to magnetic beads

Rabbit IgG antibodies (Sigma) were coupled in-house on BcMag™ Epoxy-activated Magnetic beads (Bioclone) as previously described (Hamperl et al., 2014).

Cross-linking and compatibleness with pull-down

Cross-linking of *Arabidopsis* cell suspension cultures was obtained using formaldehyde 3 days after subculturing. Cultures were treated for 10 min with 0,75% formaldehyde followed by quenching the cross-linking reaction with 250 mM glycine for 10 min. Cross-linked cells were harvested as previously described (Van Leene et al., 2007). To compare the pull-down efficiency with and without cross-linking, 5 g of harvested cells were used for ChAP analysis as described below.

Chromatin affinity purification (ChAP)

20 g cross-linked and/or non-cross-linked cells were ground to homogeneity in liquid nitrogen. Four volumes of ChAP-MS buffer without detergents (50 mM HEPES pH7.5, 150 mM NaCL, 0,5 mM EDTA, 1 µM trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E64), 1 mM PMSF, 0.1 mM sodium vanadate, 1 mM NaF, complete ultra EDTA free tablet (Roche)) was added followed by mixing with an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC) at 4°C. Detergents (1% Triton X-100, 0.1% NaDoc, 0.1% SDS) were added and the crude extracts were incubated on a horizontal shaker for 30 min at 4°C. Extracts were sonicated for 4,5 min (10 sec ON, 20 sec OFF), and supernatants were separated from cell debris by centrifugation at 2000g for 10 min at 4°C. The extracts were passed through a GF-prefilter and 0.45-µm filter (Sartorius AG), and protein concentrations were determined using the Bradford assay. The total extracts were incubated with 800 µL in-house prepared magnetic IgG beads (in case of GS^{rhino}) or 400 µL GFP-TRAP agarose beads (Chromotek) (in case of GS^{yellow}) overnight at 4°C under gentle rotation. Beads were washed with 10 mL of ChAP-MS buffer, 2x 750 µL low salt buffer, 2x 750 µL high salt buffer and 2x 750 µL LiCl buffer. A fraction of the beads was kept aside for DNA purification. **DNA purification:** For reverse cross-linking, 10 volumes Dex-buffer (10 mM Tris pH8, 1mM EDTA, 0.5M NaCl, 1% SDS) + 0.5 µl RNAse A was added to the beads, while input samples were reverse cross-

linked by adding 3 volumes of water, 40 µL 5M NaCl and 0,6 µL RNase at 250 µL crude extract. Reverse cross-linking went on overnight at 65°C. Reverse cross-linked samples were centrifuged for 5 min at 14000 rpm, followed by 50 µg proteinase K treatment of 2 hours at 50°C. DNA was purified by the fenol/chloroform/IAA purification strategy and concentrated using the Qiaquick PCR purification Kit (Qiagen). Purified DNA was analyzed with qRT-PCR (in case of ribosomal DNA) using the LightCycler 480 Real-Time SYBR Green PCR System (Roche). Used primers are listed in supplementary table 1, including the housekeeping gene ACTIN 2 (AT3G18780). Purified DNA from telomere targeting was subjected to telomere amplification as described below. To check sonication, purified DNA of input samples were loaded on an 1,2% agarose gel. **Protein purification:** The other fraction of magnetic IgG beads was subjected to 0.5 N ammonium hydroxide / 0.5 mM EDTA for 30 min at room temperature to elute proteins. The eluates were lyophilized, and the protein pellet was dissolved in 30 µL 1x NuPAGE buffer (ThermoFisher Scientific). The fraction of GFP-TRAP agarose beads was boiled for 30 min in 2x sample buffer to elute the proteins. Eluted proteins were separated from the beads on a Mobicol F colum (Mobi Tec). Proteins were precipitated in ice cold ethanol overnight at -80°C, and the pellet was resolved in 30 µL 1x NuPAGE buffer (ThermoFisher Scientific). Protein samples were loaded on a Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad), followed by blotting onto a PVDF membrane using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). GS^{rhino}-tagged dCas9 was detected with Peroxidase Anti-Peroxidase Soluble Complex antibody (Sigma), while GS^{yellow} fusions were detected with polyclonal anti-GFP (AB290, Abcam).

Amplification of telomere sequences

Amplification of telomere sequences is based on the monochrome multiplex quantitative PCR method (MMQPCR) (Cawthon, 2009). PCRs were executed with SYBR Green PCR mix (Roche) or Hotfire polymerase mix (Solis Biodyne). In a first step the polymerase is activated at high temperature level (95°C) for 15 min. In a subsequent stage, two cycles of 15 sec at 94°C and 15 sec at 49°C allows specific annealing of the telomere primers with telomere sequences, and prevents primer dimer formation. Hereafter, a standard PCR reaction is followed with 32 cycles of 15 sec at 94°C, 10 sec at 62°C and 15 sec at 74°C. Amplified PCR fragments were loaded on an 1,2% agarose gel.

References

Bae S, Park J, Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30:** 1473-1475

Beilstein MA, Renfrew KB, Song X, Shakirov EV, Zanis MJ, Shippen DE (2015) Evolution of the Telomere-Associated Protein POT1a in Arabidopsis thaliana Is Characterized by Positive Selection to Reinforce Protein-Protein Interaction. Mol Biol Evol **32:** 1329-1341

Besbrugge N, Van Leene J, Eeckhout D, Cannoot B, Kulkarni SR, De Winne N, Persiau G, Van De Slijke E, Bontinck M, Aesaert S, Impens F, Gevaert K, Van Damme D, Van Lijsebettens M, Inze D, Vandepoele K, Nelissen H, De Jaeger G (2018) GS(yellow), a Multifaceted Tag for Functional Protein Analysis in Monocot and Dicot Plants. Plant Physiol **177:** 447-464

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Byrum SD, Raman A, Taverna SD, Tackett AJ (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep **2:** 198-205

Byrum SD, Taverna SD, Tackett AJ (2015) Purification of specific chromatin loci for proteomic analysis. Methods Mol Biol **1228:** 83-92

Cawthon RM (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic Acids Res **37:** e21

Chari R, Mali P, Moosburner M, Church GM (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods **12:** 823-826

Earley KW, Pontvianne F, Wierzbicki AT, Blevins T, Tucker S, Costa-Nunes P, Pontes O, Pikaard CS (2010) Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes Dev **24:** 1119-1132

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J **79:** 348-359

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y, Zhu JK (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res **23:** 1229-1232

Fitzgerald MS, Riha K, Gao F, Ren S, McKnight TD, Shippen DE (1999) Disruption of the telomerase catalytic subunit gene from Arabidopsis inactivates telomerase and leads to a slow loss of telomeric DNA. Proc Natl Acad Sci U S A **96:** 14813-14818

Fujii H, Fujita T (2015) Isolation of Specific Genomic Regions and Identification of Their Associated Molecules by Engineered DNA-Binding Molecule-Mediated Chromatin Immunoprecipitation (enChIP) Using the CRISPR System and TAL Proteins. Int J Mol Sci **16:** 21802-21812

Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, Fujii H (2013) Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci Rep **3:** 3171

Fujita T, Fujii H (2013) Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun **439:** 132-136

Fujita T, Fujii H (2013) Locus-specific biochemical epigenetics/chromatin biochemistry by insertional chromatin immunoprecipitation. ISRN Biochem **2013:** 913273

Fujita T, Fujii H (2014) Identification of proteins associated with an IFNgamma-responsive promoter by a retroviral expression system for enChIP using CRISPR. PLoS One **9:** e103084

Fujita T, Yuno M, Okuzaki D, Ohki R, Fujii H (2015) Identification of non-coding RNAs associated with telomeres using a combination of enChIP and RNA sequencing. PLoS One **10:** e0123387

Gao XD, Rodriguez TC, Sontheimer EJ (2019) Adapting dCas9-APEX2 for subnuclear proteomic profiling. Methods Enzymol **616:** 365-383

Gao XD, Tu LC, Mir A, Rodriguez T, Ding Y, Leszyk J, Dekker J, Shaffer SA, Zhu LJ, Wolfe SA, Sontheimer EJ (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods

Gaudino RJ, Pikaard CS (1997) Cytokinin induction of RNA polymerase I transcription in Arabidopsis thaliana. J Biol Chem **272:** 6799-6804

Hamperl S, Brown CR, Perez-Fernandez J, Huber K, Wittner M, Babl V, Stockl U, Boeger H, Tschochner H, Milkereit P, Griesenbeck J (2014) Purification of specific chromatin domains from single-copy gene loci in Saccharomyces cerevisiae. Methods Mol Biol **1094:** 329-341

Hoshino A, Fujii H (2009) Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. J Biosci Bioeng **108:** 446-449

Huffman KE, Levene SD, Tesmer VM, Shay JW, Wright WE (2000) Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. J Biol Chem **275:** 19719-19722

Hwang MG, Chung IK, Kang BG, Cho MH (2001) Sequence-specific binding property of Arabidopsis thaliana telomeric DNA binding protein 1 (AtTBP1). FEBS Lett **503:** 35-40

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337:** 816-821

Jodczyk S, Pearson JF, Aitchison A, Miller AL, Hampton MB, Kennedy MA (2015) Telomere length measurement on the Roche LightCycler 480 Platform. Genet Test Mol Biomarkers **19:** 63-68

Koukalova B, Fojtova M, Lim KY, Fulnecek J, Leitch AR, Kovarik A (2005) Dedifferentiation of tobacco cells is associated with ribosomal RNA gene hypomethylation, increased transcription, and chromatin alterations. Plant Physiol **139:** 275-286

Lawrence RJ, Pikaard CS (2004) Chromatin turn ons and turn offs of ribosomal RNA genes. Cell Cycle **3:** 880-883

Layat E, Saez-Vasquez J, Tourmente S (2012) Regulation of Pol I-transcribed 45S rDNA and Pol III-transcribed 5S rDNA in Arabidopsis. Plant Cell Physiol **53:** 267-276

Li H, Tsang CK, Watkins M, Bertram PG, Zheng XF (2006) Nutrient regulates Tor1 nuclear localization and association with rDNA promoter. Nature **442:** 1058-1061

Liang G, Zhang H, Lou D, Yu D (2016) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci Rep **6:** 21451

Liu X, Zhang Y, Chen Y, Li M, Shao Z, Zhang MQ, Xu J (2018) CAPTURE: In Situ Analysis of Chromatin Composition of Endogenous Genomic Loci by Biotinylated dCas9. Curr Protoc Mol Biol**:** e64

Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, Dickerson KE, Xie S, Hon GC, Xuan Z, Zhang MQ, Shao Z, Xu J (2017) In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell **170:** 1028-1043 e1019

Lundblad V, Blackburn EH (1993) An alternative pathway for yeast telomere maintenance rescues est1- senescence. Cell **73:** 347-360

Lydall D (2003) Hiding at the ends of yeast chromosomes: telomeres, nucleases and checkpoint pathways. J Cell Sci **116:** 4057-4065

Mair A, Xu SL, Branon TC, Ting AY, Bergmann DC (2019) Proximity labeling of protein complexes and cell type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife **8**

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol **31:** 833-838

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science **339:** 823-826

Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu JK (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant **6:** 2008-2011

Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat Methods

Myers SA, Wright J, Zhang F, Carr SA (2017) CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus. Molecular & Cellular Proteomics **16:** S63-S63

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol **31:** 691-693

Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, Van Lijsebettens M, Goormachtig S, Vandepoele K, Van Leene J, Muszynski M, Gevaert K, Inze D, De Jaeger G (2015) Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf. Plant Cell **27:** 1605-1619

Nelson AD, Shippen DE (2012) Surprises from the chromosome front: lessons from Arabidopsis on telomeres and telomerase. Cold Spring Harb Symp Quant Biol **77:** 7-15

Prochazkova Schrumpfova P, Fojtova M, Fajkus J (2019) Telomeres in Plants and Humans: Not So Different, Not So Similar. Cells **8**

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell **152:** 1173-1183

Ren M, Qiu S, Venglat P, Xiang D, Feng L, Selvaraj G, Datla R (2011) Target of rapamycin regulates development and ribosomal RNA expression through kinase domain in Arabidopsis. Plant Physiol **155:** 1367-1382

Rowland EE (2017) PROTEOLYTIC MATURATION AND PROTEIN DEGRADATION IN ARABIDOPSIS THALIANA CHLOROPLASTS. Cornell University

Saez-Vasquez J, Delseny M (2019) Ribosome Biogenesis in Plants: From Functional 45S Ribosomal DNA Organization to Ribosome Assembly Factors. Plant Cell **31:** 1945-1967

Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H (2016) Determination of local chromatin composition by CasID. Nucleus **7:** 476-484

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, Martens L, Witters E, Gevaert K, De Jaeger G (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc **10:** 169-187

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, Henderickx K, De Vijlder T, Abdelkrim A, Pharazyn A, Van Onckelen H, Inze D, Witters E, De Jaeger G (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics **6:** 1226-1238

Vaquero-Sedas MI, Vega-Palas MA (2014) Determination of Arabidopsis thaliana telomere length by PCR. Sci Rep **4:** 5540

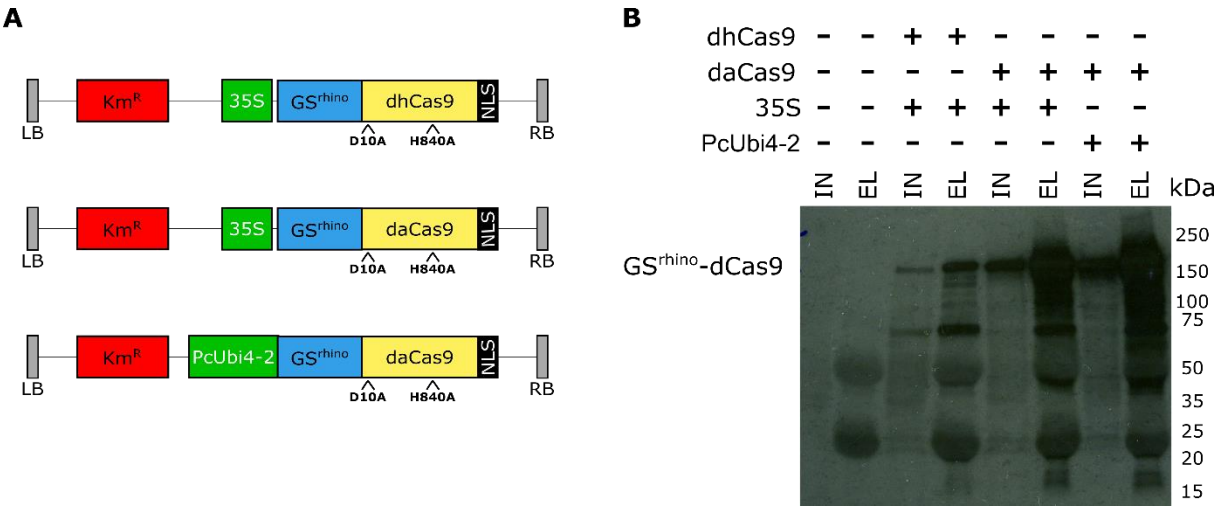
Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Wahls WP, Taverna SD, Raney KD, Tackett AJ (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics **9:** 1207-1211

Wei Y, Tsang CK, Zheng XF (2009) Mechanisms of regulation of RNA polymerase III-dependent transcription by TORC1. EMBO J **28:** 2220-2230

Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol Plant **6:** 1975-1983

Zhou X, Jacobs TB, Xue LJ, Harding SA, Tsai CJ (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate:CoA ligase specificity and redundancy. New Phytol **208:** 298-301

Supplementary data



Supplementary figure 1: GS^{rhino}-dCas9 protein accumulation under different promoters. **A)** T-DNA constructs overexpressing human codon optimized dead Cas9 (dhCas9) or *Arabidopsis* codon optimized dead Cas9 (daCas9) under the 35S promoter or Ubiquitin4–2 promoter from *Petroselinum crispum* (PcUbi4-2) **B)** Immunoblot analysis of wild type (PSB-D) and tagged human (dhCas9) or *Arabidopsis* (daCas9) codon optimized Cas9 overexpressing cell suspension culture before (IN, 100 µg total protein) and after (EL, 25 µL total eluate) pull-down using Peroxidase Anti-Peroxidase Soluble Complex antibody. daCas9 expression was under control of the 35S promoter or constitutive Ubiquitin4–2 promoter from *Petroselinum crispum* (PcUbi4-2).

Supplementary table 1: Primers used for cloning, sequencing and qPCR analysis

Name	Sequence	Application
hCas9 D10A F	gtactccattgggctcgctatcggcacaa	Site directed mutagenesis
hCas9 D10A R	gagcccaatggagtacttcttgtccatggtgaa	Site directed mutagenesis
hCas9 H840A F	cgactacgacgtggatgctatcgtgcccc	Site directed mutagenesis
hCas9 H840A R	atccacgtcgtagtcggagagccgattgat	Site directed mutagenesis
dhCas9 int1	tggaggagtccttttggtg	Sequencing
dhCas9 int2	aaagacacctacgatgatga	Sequencing
dhCas9 int3	agataacagggaaaagattg	Sequencing
dhCas9 int4	ttaaagacaaggacttcctg	Sequencing
dhCas9 int5	caaactacccagaagggaca	Sequencing
dhCas9 int6	tctcgattcacgcatgaaca	Sequencing
dhCas9 int7	agaccggaggcttctccaag	Sequencing
dhCas9 int8	gagatcatcgagcaaataag	Sequencing
attB1-NLS-dCas9 F	ggggacaagtttgtacaaaaaagcaggctccatggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccatggacaagaagtactccattggg	Gateway cloning
attB2-dhCas9-NoStop R	ggggaccactttgtacaagaaagctgggtccaccttcctcttcttcttgggg	Gateway cloning
Oligo telomere F	tgattggggtttagggtttagggtttg	Oligo annealing
Oligo telomere R	aaaacaaaccctaaaccctaaacccca	Oligo annealing
Oligo 5SgRNA1 F	tgattgcaacacgaggacttcccgggg	Oligo annealing
Oligo 5SgRNA1 R	aaaaccccgggaagtcctcgtgttgca	Oligo annealing
Oligo 5SgRNA2 F	tgattgtaaaagagggatgcaacacgg	Oligo annealing
Oligo 5SgRNA2 R	aaaaccgtgttgcatccctcttttaca	Oligo annealing
Oligo 45SgRNA F	tgattggagggagtctgggcagtccgg	Oligo annealing
Oligo 45SgRNA R	aaaaccggactgcccagactccctcca	Oligo annealing
5S rDNA primer set 2 F	ggatgcgatcataccagc	qPCR
5S rDNA primer set 2 R	gagggatgcaacacgagg	qPCR
5S rDNA primer set 1 F	ataccagcactaatgcaccg	qPCR
5S rDNA primer set 1 R	aaagagggatgcaacacgag	qPCR
45S rDNA F	tcggatatcgacacgaggaa	qPCR

45S rDNA R	cggaaaagtcgccgaaaaag	qPCR
5SgRNA1 off Ch5 F	gacgtatgcagtgggctagt	qPCR
5SgRNA1 off Ch5 R	caatgaaccatttcttcttcccctc	qPCR
5SgRNA1 off Ch2 F	tgggatccgttgcattcgtt	qPCR
5SgRNA1 off Ch2 R	ccacagcgaatggttcctaga	qPCR
5SgRNA2 off Ch2(2) F	tgggatccgttgcattcgtt	qPCR
5SgRNA2 off Ch2(2) R	ccacagcgaatggttcctaga	qPCR
5SgRNA2 off Ch2(4) F	gctgtggattagcttccccg	qPCR
5SgRNA2 off Ch2(4) R	ttgagttcttctgctctgcttct	qPCR
45SgRNA off Ch1 F	cgaggaatgaccgatcgac	qPCR
45SgRNA off Ch1 R	tcaaataaccgatcgccaga	qPCR
45SgRNA off Ch5 F	tcagtagtattgcctccagcacag	qPCR
45SgRNA off Ch5 R	tggctttgtggtgggattctgg	qPCR

Chapter 5

A hunt down for mitotic APC substrates.

Caroline Matthijs[1,2], Jelle Van Leene[1,2], Astrid Gadeyne[1,2], Geert Persia[u1,2], Nancy De Winne[1,2], Eveline Van De Slijke[1,2], Dominique Eeckhout[1,2], Francis Impens[3], Geert De Jaeger[1,2]

Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

VIB Proteomics Core, 9000 Ghent, Belgium[3]

Author contributions

CM, JVL, AG and GDJ designed the research; NDW and EVDS performed cloning and transformation experiments; CM and GP performed the research; mass spectrometry was performed by FI and mass spectrometry data was analyzed by DE; CM analyzed the research data and wrote the chapter; JVL and GDJ supervised and complemented the writing.

Abstract

Controlled degradation of proteins using the ubiquitin/26S proteasome pathway allows for removal of misfolded and non-functional proteins as well as maintaining homeostasis by controlling the abundancy of regulatory proteins. In plants, different E3 ligases, which ubiquitinate target proteins, exist of which the anaphase promoting complex (APC) is the most intricate one. While regulation of this E3-ligase is being clarified, identification of substrates is lacking behind in plants. In this chapter we applied a pull-down strategy on propyzamide synchronized *Arabidopsis* cell suspension culture to reveal common and specific interactions for the three APC-specific CCS52 co-activators during mitosis. Next to the pull-down of an almost complete APC complex together with known regulators, we also retrieved known and putative new APC substrates. We also applied biotin-based proximity labeling to validate and further extend the CCS52B interactome. Our data suggests a mitotic role for CCS52B, and provides an interesting amount of putative mitotic APC substrates which need to be further validated.

Introduction

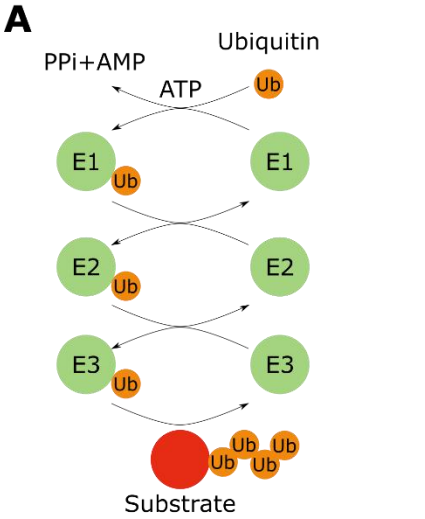
One important aspect of a plant's life is the controlled degradation of proteins. This regulated proteolysis allows for removal of misfolded and non-functional proteins as well as maintaining homeostasis by controlling the abundancy of regulatory proteins. The most common proteolysis system in eukaryotes is the ubiquitin/26S proteasome pathway (Sullivan et al., 2003; Smalle and Vierstra, 2004). This pathway consists of different players allowing the ubiquitination of target proteins and the subsequent degradation by the 26S proteasome. The key component is the reusable, small, 76-amino acid protein ubiquitin (Ub). In a first step, Ub is bound and activated by the Ub-activating enzyme (E1) using adenosine triphosphate (ATP). Next, Ub is transferred to an Ub-conjugating enzyme (E2) that together with an Ub-protein ligase (E3) covalently binds the Ub to a lysine residue of a target protein (figure 1A). Depending on the target and the E2/E3 complex, different Ub patterns can be formed. Monoubiquitination frequently directs substrates to the lysosome/vacuole, while Lys48-linked Ub polymers are preferred for 26S proteasomal degradation.

In *Arabidopsis* two E1 isoforms, 37 E2 (or UBC) genes and more than 1300 E3 genes are expressed. E3 controls the specificity of the ubiquitination cascade by recognizing and directly binding the target protein. The large number of E3 genes in plants suggest that the amount of targets may number in the thousands. E3s are classified in 4 types (HECT, RING/U-box, SCF and APC) based on subunit composition and mode of action (Sullivan et al., 2003; Smalle and Vierstra, 2004). The APC (anaphase-promoting complex) is the most intricate E3 type, consisting of a core complex of 14 subunits in plants and yeast or 19 subunits in human cells, and interchangeable co-activator subunits (figure 1B-C) (Eloy et al., 2015; Alfieri et al., 2017; Lorenzo-Orts et al., 2019). This ubiquitin-protein ligase mainly functions during cell division but also regulates other cellular processes like cell differentiation, genome stability, energy metabolism, cell death, autophagy as well as carcinogenesis (Zhou et al., 2016).

The complex can be subdivided in three parts (figure 1B): a platform, a structural module and the catalytic core. The largest subunit of the APC (APC1) forms, together with APC4 and APC5, a platform that links the structural module with the catalytic core (da Fonseca et al., 2011). The structural module includes APC6, APC7, APC8 and APC3. For the latter two genes are present in the *Arabidopsis* genome (APC3a/CDC27a and APC3b/HOBBIT) (Blilou et al., 2002). The catalytic core consists of APC2 and APC11, and together they are sufficient to catalyze non-specific ubiquitination reactions *in vitro* (Gmachl et al., 2000; Tang et al., 2001). Substrate specificity is obtained by the subunit APC10 and the co-activators CELL DIVISION CYCLE 20 (CDC20) and Cell Cycle Switch 52 (CCS52, plant ortholog of human

CDC20 Homolog 1 (CDH1)) which are located within the central cavity of the APC (da Fonseca et al., 2011). In yeast and mammals, the APC also contains two extra subunits, APC13 and APC15, for which homologs have been found in plants (Schwickart et al., 2004; Uzunova et al., 2012). However, these subunits have not been co-purified by tandem affinity purification (TAP) of the *Arabidopsis* APC complex, while all other subunits have been shown to interact with each other (Van Leene et al., 2010). Mutation studies of different APC subunits in *Arabidopsis* resulted in non-viable plants, an inactive E3 ligase, the accumulation of cyclin substrates and mitotic arrest, showing the essential role of the APC complex during early plant development (Eloy et al., 2011; Heyman and De Veylder, 2012; Wang et al., 2012; Wang et al., 2013; Eloy et al., 2015).

A correct regulation of the APC is crucial and is carried out by transcriptional regulation, posttranslational modifications, inhibitor and co-activator proteins. Three plant specific APC inhibitors have been reported, known as Ultraviolet-B-Insensitive 4 (UVI4) (Heyman et al., 2011), Omission of second division 1 (OSD1) (Cromer et al., 2012) and SAMBA (Eloy et al., 2012). Although, SAMBA shows some characteristics of a co-activator, targeting the degradation of CycA2;3 (Eloy et al., 2012; Heyman and De Veylder, 2012). Inhibition of APC activity occurs by interacting and blocking the APC co-activators CDC20 and CCS52 (Iwata et al., 2011). The co-activators bind, activate and provide substrate specificity of the APC complex. In *Arabidopsis* five genes encode for putative CDC20 proteins of which two are functional (CDC20.1 and CDC20.2) (Kevei et al., 2011). For CCS52 three genes are present in the *Arabidopsis* genome (CCS52A1, CCS52A2 and CCS52B). The co-activators have three characteristic protein domains essential for their function, including a N-terminal C-box, a C-terminal IR-tail and a WD40 domain (figure 1D) (Matyskiela and Morgan, 2009). The IR-tail binds to the APC3 subunit while the C-box interacts with the APC8 subunit, both ensuring stable binding of the co-activators (Yamano, 2019). Via the WD40 domain the co-activator can bind APC target proteins by recognizing a destruction motif (see further). Expression of the co-activators is cell cycle phase dependent. The two CDC20 genes are expressed from early G2-phase until the M-phase exit, while CCS52B is expressed from G2/M- to M-Phase and CCS52A1 and CCS52A2 proteins are present from late M-phase until early G2-phase (Menges et al., 2003; Fulop et al., 2005). This expression profile suggests consecutive actions of the different co-activators during the plant cell cycle and thus a transcriptional regulation of the APC complex. In contrast, most APC core subunits are constitutively expressed, with exception for the APC3 a and b subunits, showing an increase in expression during S- and G2-phase (Heyman and De Veylder, 2012).



C

	Human	Yeast	Arabidopsis
Core complex	APC1	APC1	APC1
	APC2	APC2	APC2
	APC3A APC3B	Cdc27	APC3A APC3B
	APC4	APC4	APC4
	APC5	APC5	APC5
	APC6A APC6B	Cdc16	APC6
	APC7A APC7B		APC7
	APC8A APC8B	Cdc23	APC8
		APC9	
	APC10	Doc1	APC10
	APC11	APC11	APC11
	APC12A APC12B	Cdc26	APC12*
	APC13	Swm1	APC13
	APC15	Mnd2	APC15
	APC16	APC16	
Co-activators	CDC20	CDC20	CDC20.1 CDC20.2
	Cdh1	Cdh1	CCS52A1 CCS52A2 CCS52B

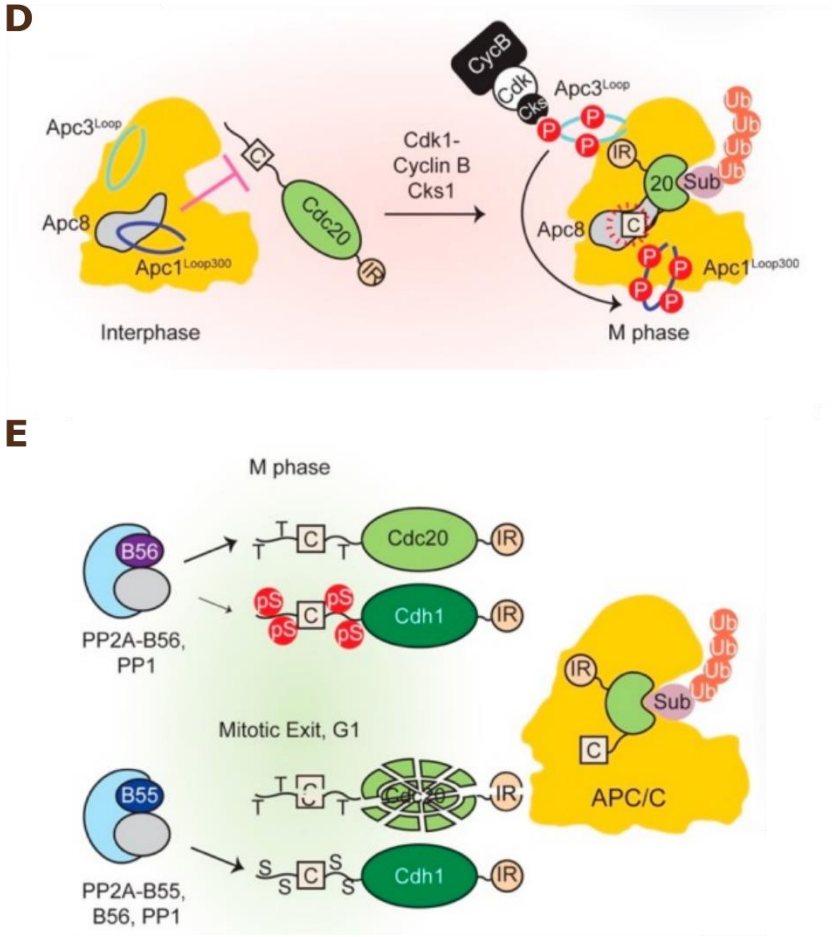


Figure 1: Function, structure and regulation of the E3 ubiquitin ligase anaphase promoting complex (APC). A) Scheme of the ubiquitination reaction by the Ub-activating enzyme (E1), the Ub-conjugating enzyme (E2) and Ub-protein ligase (E3). **B)** Schematic representation of the plant APC complex. Green: platform module; red: structural module; brown: catalytic and substrate recognition module; orange: activators subunit; purple: plant specific APC interactors. Figure adapted from (Eloy et al., 2015). **C)** Overview of the different core APC subunits and interchangeable co-activators known for human, yeast and Arabidopsis. *APC12 has recently been discovered in Arabidopsis (Lorenzo-Orts et al., 2019). **D)** The Apc1 disordered loop domain ($Apc1^{Loop300}$) occludes the C-box binding site within Apc8 and prevents Cdc20 loading. Recruitment of mitotic kinases by phosphorylation of the disordered APC3 loop domain ($Apc3^{Loop}$) leads to $Apc1^{Loop300}$ phosphorylation and Cdc20 loading. Figure adopted from (Kataria and Yamano, 2019). **E)** In mitosis, PP2A-B56 and PP1 dephosphorylate the threonine-rich Cdc20 N-terminus to promote its interaction with the APC/C. During mitotic exit and in G1, PP2A-B55 (and other mitotic exit phosphatases) dephosphorylate serine-rich Cdh1 N-terminus and promotes co-activator exchange and Cdc20 degradation. Figure adopted from (Kataria and Yamano, 2019).

Before co-activators can interact with the APC complex, activation by phosphorylation at onset of mitosis is needed. This is carried out by a Cyclin-CDK-CKS complex which phosphorylates in different steps the disordered loop domains of APC3 and APC1, allowing co-activator association (figure 1D) (Patra and Dunphy, 1998; Zhang et al., 2016; Kataria and Yamano, 2019; Yamano, 2019). During interphase, the disordered loop domain of APC1 blocks the C-box binding site on APC8 and thus inhibits co-activator binding. At onset of mitosis, CDK-CycB-CKS will phosphorylate the disordered loop domains of the APC3 subunits and will associate with these PTMs. Subsequently, phosphorylation of the APC1 disordered loop domain is carried out, dislocating it from the C-box binding site and allowing co-activator association (figure 1D). Next to phosphorylation of the disordered loops by Cyclin-CDK-CKS, several other phosphorylation sites on the APC complex have been reported but their function remains elusive (Kraft et al., 2003; Yamano, 2019). An extra level of APC activity regulation is executed by different phosphatases, which dephosphorylate the N-terminal tail of the co-activators (figure 1E) (Kataria and Yamano, 2019). At onset of mitosis, the B56 phosphatase family is responsible for the specific dephosphorylation of the CDC20 N-terminal tail in human cells, allowing CDC20 association with the APC complex. Later on, at onset of anaphase, the B55 phosphatase family will dephosphorylate the CDH1 N-terminal tail, stimulating co-activator exchange of CDC20 to CDH1 and subsequent CDC20 degradation (figure 1E).

A specific and time dependent degradation of APC substrates is essential for a normal cell cycle progression. APC substrates contain destruction motifs (or degrons) that are recognized and bound by the APC activators. The three major degrons are the destruction box (D-box, RxxLxxxxN) (Glotzer et al., 1991), the KEN-box (KENxxxN/D) (Pfleger and Kirschner, 2000) and the ABBA motif ([FILV]x[ILMVP][FHY]x[DE]) (Di Fiore et al., 2015; Davey and Morgan, 2016). In mammals, APC^{CDC20} targets Cyclin A, Cyclin B and Securin (figure 2) (Zhou et al., 2016). Degradation of these essential cell cycle regulators allows for the progression from prometaphase to onset of anaphase. By degrading Securin and cyclins, Separase is activated upon which the Cohesin protein complex is cleaved and sister chromatids come apart from each other (Gorr et al., 2005). Recently, Arabidopsis homologs for Securin have been identified, called PATRONUS 1 and PATRONUS 2 (PANS1 and PANS2) (Cromer et al., 2019). Also they contain degrons and interact with the APC. A Y2H screen confirmed the interaction of CDC20.1 and CDC20.2 with specific plant cyclins (Kevei et al., 2011). Although all mitotic cyclins contain D-box sequences, not all of them are targeted by APC^{CDC20}. Probably APC^{CDC20} has a specificity towards a subset of variant D-box sequences while other D-box variants are targeted by APC^{CCS52}, suggesting that variation in D-box sequence and the different CDC20 and CCS52 isoforms ensure for

the degradation of specific mitotic cyclins at a given stage during cell cycle (Kevei et al., 2011; Davey and Morgan, 2016).

At onset of anaphase in mammals, CDH1 is activated and is responsible for the further progression through mitosis and inhibition of Cyclin-CDK activity during G1 phase. Known APC^{CDH1} substrates are B-type cyclins (CycB), Aurora, CDC20 and many more (figure. 2) (Kramer et al., 2000; Zhou et al., 2016). While the list of CDH1 targets is comprehensive, the amount of known substrates for the plant CDH1 ortholog CCS52 is limited (Heyman and De Veylder, 2012). Stabilization of CycA2;3 has been observed in mutant ccs52a1 plants (Boudolf et al., 2009), and in 2010 a TAP experiment revealed novel interactors of the plant APC (Van Leene et al., 2010), containing at least one CCS52A2 substrate, ERF115 (Heyman et al., 2013). Both A-type CCS52's have been demonstrated to control endocycle onset in leaves and roots, and control stem cell maintenance (Lammens et al., 2008; Boudolf et al., 2009; Vanstraelen et al., 2009). Furthermore, localization studies have shown the association of CCS52A2 with the preprophase band (PPB) at onset of mitosis (Boruc et al., 2010). However, the function of CCS52B during cell cycle remains unclear.

Based on the expression profile of CCS52B, which exhibits a M-phase peak (Menges et al., 2003), one can postulate that this APC co-activator plays a main role during mitosis and may fulfill the mitotic role of CDH1 in *Arabidopsis*. Further evidence for a M-phase dependent function of CCS52B was provided by Yang et al. (2017), who observed expression of CDC20 and CCS52B during prophase and sequestration of the corresponding mRNA in the nucleus until nuclear envelope breakdown (NEB) at prometaphase (Yang et al., 2017). While CDC20 and CCS52B mRNA levels peaked at prophase, a decrease in CDC20 mRNA was observed at the end of mitosis, whereas the high CCS52B mRNA level remain until cytokinesis. This again illustrates the consecutive actions of the different APC co-activators. The nuclear sequestration of CDC20 and CCS52B mRNA prevents early protein translation and is as such an extra mechanism to tightly control APC during cell cycle. Only after NEB, the high mRNA levels of CDC20 and CCS52B are translated and this allows for a rapid and efficient activation of the APC (Yang et al., 2017). Furthermore, TAP analysis on SAMBA in maize leafs showed a significant enrichment of CCS52B in dividing cells compared to expanding cells (data not published, Bontinck M.).

Interaction of the CCS52 co-activators with different mitotic cyclins has been demonstrated with Y2H screens, bimolecular fluorescence complementation (BiFC) and co-immunoprecipitations (Fulop et al., 2005; Boruc et al., 2010). Similar to CDC20, there is selectivity for specific cyclins. For example, a strong interaction is observed for CycA3;4 with CCS52A1 and CCS52B, but binding is weaker with CCS52A2. Other

mitotic cyclins interacting with CCS52B are CycA1;1, CycB1;1 and CycB1;2. However, interaction of CCS52B with CycA1;2 was significant weaker than with CCS52A1 and CCS52A2 (Fulop et al., 2005). Localization studies of these mitotic cyclins during cell division revealed protein degradation around metaphase or anaphase, further evidence for these cyclins to be targets of the APC (Boruc et al., 2010).

Here, we perform our pull-down strategy (AP-MS) on propyzamide synchronized cell culture for all three CCS52 co-activators to further supplement the list of APC^{CCS52} substrates during cell division in *Arabidopsis*. We present a list of known and unknown interactions, including specific interactors for the three co-activators. By identifying new specific APC substrates, we get closer to unravel the role of APC during cell division in plants.

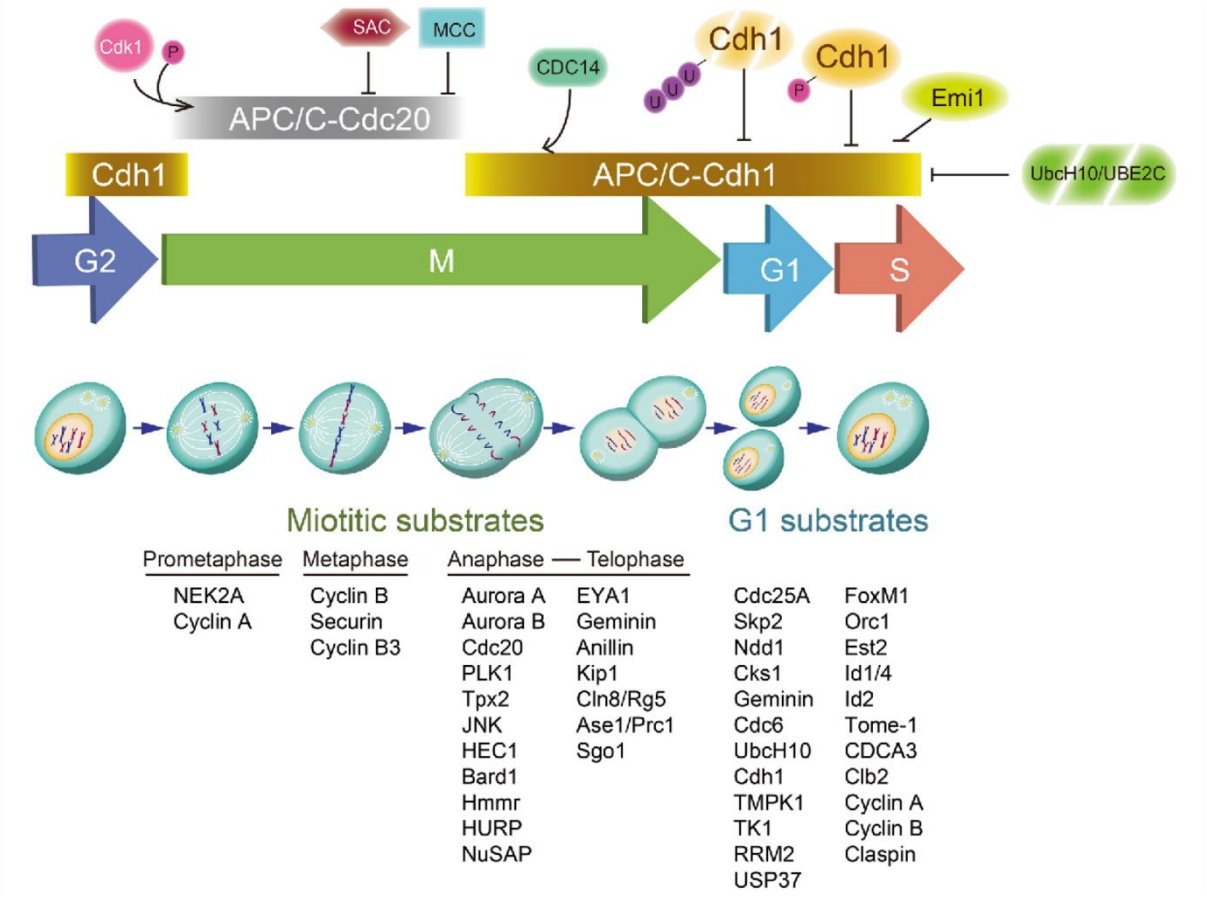


Figure 2: Cell cycle regulation by APC in mammals. Figure adopted from (Zhou et al., 2016)

Results

Propyzamide enriches for a mitotic cell suspension culture.

To identify CCS52 specific substrates during cell division it is important to perform our AP-MS at the time point during mitosis where APC^{CCS52} is activated. We choose to analyze the metaphase-anaphase transition, the time-point when CDC20 is exchanged for CDH1 and different CDH1 targets are being degraded (Zhou et al., 2016). AP-MS is performed on *Arabidopsis* cell suspension cultures (PSB-D), a biological system that allows efficient cell cycle synchronization through for example sucrose starvation and repletion or aphidicolin treatment (Menges and Murray, 2002). Different strategies and chemicals can be applied to study cell cycle via synchronization (Planchais et al., 2000). Blocking of mitosis can be achieved by disturbing the microtubule polymerization and formation of the mitotic spindle, resulting in a metaphase arrest. Propyzamide, a herbicide, is one of these anti-tubulin drugs that has been repeatedly used to induce mitotic arrest (Nagata et al., 1992; Nagata and Kumagai, 1999; Suzuki et al., 2005). The advantage of this drug is that it is reversible as the mitotic arrest is released and metaphase/anaphase transition resumes when the drug is removed. However, it is crucial to reduce the propyzamide treatment time to a few hours (less than 14 hours) to prevent the formation of abnormal cell division figures and micronuclei (Verhoeven et al., 1990; Planchais et al., 2000).

To be able to follow synchronization efficiency of our PSB-D cell culture, we transformed the cells with a microtubule reporter consisting of the microtubule-binding domain (MBD) of MICROTUBULE-ASSOCIATED PROTEIN 4 (MAP4) fused to GFP and driven by the constitutive 35S promoter (Marc et al., 1998). In this way microtubule dynamics during mitosis and the formation of the mitotic spindle can be visualized easily by live-imaging under a fluorescence microscope (figure 3A). We could visualize the PPB, the centralization of the nucleus and the accumulation of microtubules around the nuclear envelope during prophase; the NEB accompanied by the disappearance of the PPB, the invasion of microtubule at the space of the former nucleus and the formation of the mitotic spindle during metaphase; the depolymerization of spindle microtubule at anaphase and formation of the phragmoplast during telophase. Addition of propyzamide did not affect the prophase nor the NEB, since we observed a PPB and subsequent disappearance of it accompanied with the invasion of microtubule at the space of the former nucleus. But propyzamide clearly altered the formation of the mitotic spindle during metaphase resulting in a metaphase arrest (figure 3B).

Based on synchronization studies of tobacco BY-2 cells (Nagata et al., 1992), GFP-MAP4 culture were treated for 4 hours with 6 µM propyzamide, one day after subculturing. After release of propyzamide, the amount of cells at metaphase or anaphase (M/A index) and telophase (phragmoplast formation) were determined at different time points. Before addition of propyzamide, approximately 7% of the cells are at metaphase or anaphase. After removal of propyzamide, a clear increase in M/A mitotic figure is observed with a peak of 24% cells 10 minutes after removal (figure 3C). Also at that time point formation of phragmoplasts is observed and is increasing steep 20 min after removal of propyzamide (figure 3D).

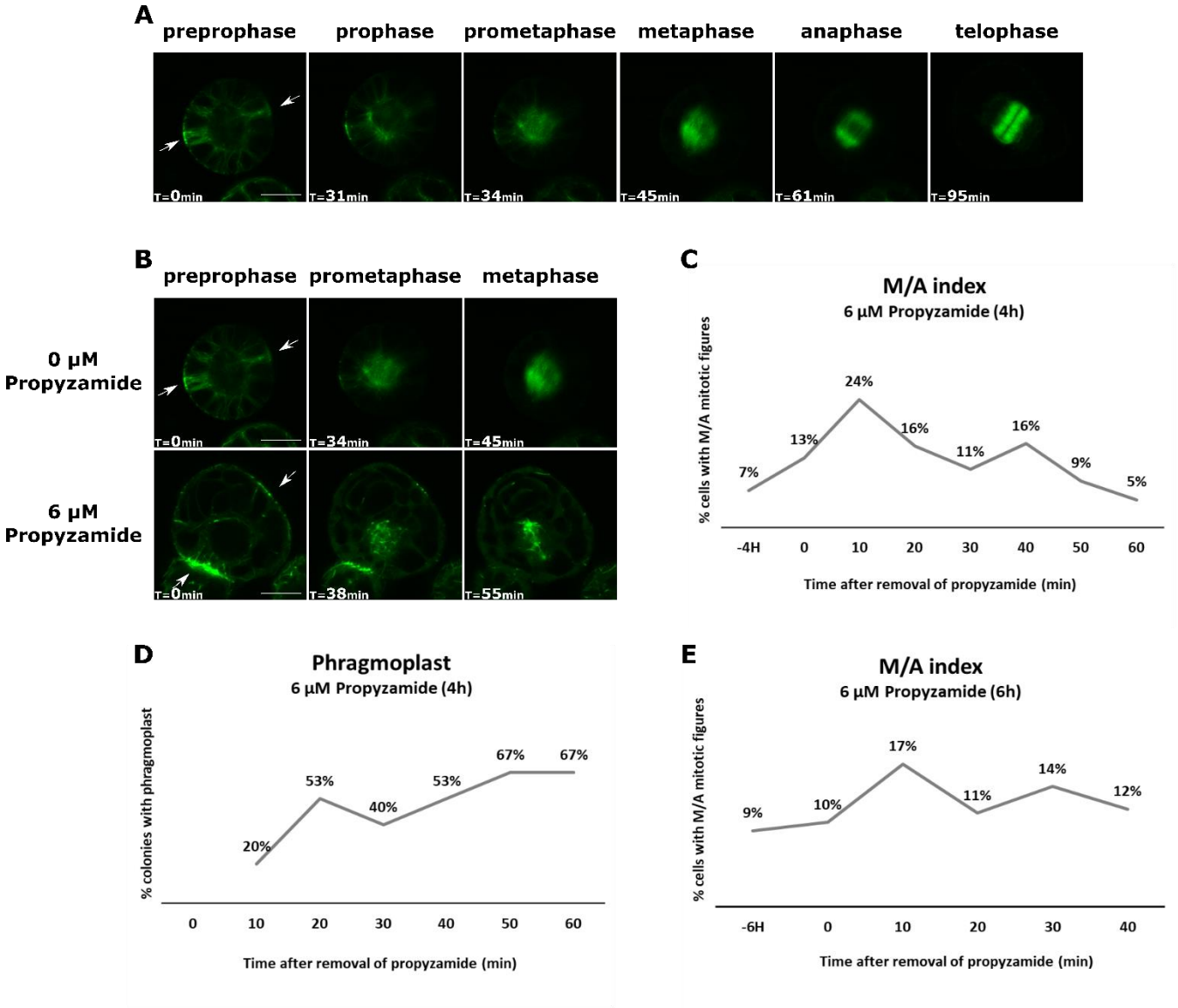


Figure 3: Propyzamide synchronization of Arabidopsis cell suspension culture. **A)** Time-lapse analysis of subcellular localization of GFP-MAP4 in PSB-D cells, visualizing the microtubule dynamics during the different mitotic phases. Arrows indicate the preprophase band. Scale bar = 10 µm **B)** Time-lapse analysis of GFP-MAP4 expressing PSB-D cells with or without propyzamide treatment. Formation of the mitotic spindle during metaphase is prevented by propyzamide. Arrows indicate the preprophase band. Scale bar = 10 µm **C)** Percentage of PSB-D cells at metaphase or anaphase (M/A index) at different time points after 4 hours of propyzamide treatment. **D)** Percentage of PSB-D colonies (+/- 15 cells) containing cells at telophase (phragmoplast formation) at different time points after 4 hours of propyzamide treatment. **E)** Percentage of PSB-D cells at metaphase or anaphase (M/A index) at different time points after 6 hours of propyzamide treatment.

These results show that with propyzamide it is possible to synchronize our PSB-D cell culture during metaphase-anaphase transition. Approximately a quarter of the cells can be synchronized after propyzamide incubation of 4 hours. A higher incubation time of 6 hours did not result in an increase of synchronized cells, but a peak 10 minutes after removal of propyzamide was still observed (figure 3E).

AP-MS on CCS52 allows pull-down of the total APC complex and specific interactors.

Propyzamide synchronization was applied on GS^{rhino}-CCS52A1, GS^{rhino}-CCS52A2 and GS^{rhino}-CCS52B overexpressing cell cultures. Cells were harvested at the arrested metaphase-anaphase transition after 4 hours propyzamide treatment (T0), and 10 minutes after removal of propyzamide. To allow for a label-free quantitative MS analysis three repeats for every time point and every bait protein were conducted. Co-purified proteins were identified on a Q Exactive (ThermoFisher Scientific) and filtered relative to a list of non-specific proteins derived from 123 AP-MS experiments, covering 13 bait groups. Proteins appearing in 3 or more bait groups were regarded as non-specific, resulting in a list of 3186 non-specific proteins (background). To prevent true interactors being filtered out because of their presence in the background list, a T-test was performed using the average spectral abundance factors (NSAF) of the identified proteins in the CCS52 pull-downs versus the corresponding average NSAF in a control set of AP-MS experiments with non-related baits (other CCS52 experiments were removed). Proteins identified with at least two peptides in at least two experiments, that were not present in the background list or showed high (at least 10-fold) and significant [-log10(p-value(T-test)) ≥10] enrichment versus the large dataset of pull-downs with non-related bait proteins, were retained. Supplementary figure 1 displays a schematic overview of the different filtering steps applied. Table 1 represents all specific and significantly enriched proteins for every time point and CCS52 bait protein.

We retrieved 57 proteins of which 20 were common for all three CCS52 co-activators. Amongst those 20 common co-purified proteins, all APC subunits and known regulators (UVI4, OSD1, SAMBA), except APC11, were retrieved with the three co-activators. When we compare these results with previous CCS52 TAP-MS purifications on non-synchronized cell cultures (Van Leene et al., 2010), it is clear that AP-MS on synchronized cell cultures is more efficient (table 1). For every CCS52 bait all APC subunits are pulled-down with AP-MS, while with TAP only a subset is retrieved. Even the small subunits APC12, APC13 and APC15 were identified, which were not observed during TAP purifications on different APC subunits (Van Leene et al., 2010). To retrieve the recently annotated APC12 subunit (Lorenzo-Orts et al., 2019), a

quantitative analysis with MaxQuant (Tyanova et al., 2016) was performed relative to a control data set of 9 AP-MS experiments on propyzamide synchronized cell cultures using the Araport11 genome annotation (Cheng et al., 2017) with APC12 (Q8H1U3) protein sequence from UniProt added. From the 57 proteins co-purified 77% interacts with CCS52B, suggesting that CCS52B is the main co-activator of the APC complex during metaphase-anaphase transition. This is in accordance with the specific mitotic upregulation of CCS52B expression (figure 3) (Menges et al., 2003; Yang et al., 2017). Scanning the *Arabidopsis* proteome with the APC/C degron repository (Davey and Morgan, 2016) reveals a total of 2025 proteins containing (putative) D-box, KEN-box and/or ABBA motifs on a total of 27655 *Arabidopsis* proteins. In our data, from the 39 APC unrelated proteins 13 contain putative major APC degrons (table 1) which is a highly significant 5.6 fold enrichment (p-value hypergeometric test = 4.8×10^{-09}). We also compared our CCS52 interactome with proteome-wide ubiquitination studies (Maor et al., 2007; Saracco et al., 2009; Kim et al., 2013; Walton et al., 2016; Aguilar-Hernandez et al., 2017; Willems et al., 2019) to reveal proteins with known ubiquitination sites. An overview is given in table 1 and the complete MS data can be retrieved from *https://floppy.psb.ugent.be/index.php/s/KnboL3zxIW3elP5*.

In addition, to correct for possible proteome changes during the propyzamide synchronization, which might lead to an altered set of non-specific binders, we also performed a quantitative MaxQuant analysis relative to AP-MS data from wild type or AUR1 and AUR2 propyzamide synchronized pull-downs, overall finding similar results. In the supplementary table 1 a comparison is made between the two filtering approaches and the complete MS data can be retrieved from *https://floppy.psb.ugent.be/index.php/s/KnboL3zxIW3elP5*. It should be noticed that the MaxQuant thresholds are stringent (FDR of 0.0001 and 0.001) and lowering these thresholds increases the overlap with the MS data retrieved after filtering relative to the large background list (data not shown).

Table 3: Proteins identified by liquid chromatography-tandem MS (LC-MS/MS) after AP-MS or TAP on the three APC co-activators CCS52A1, CCS52A2 and CCS52B. Three AP-MS analyzes for every bait after 4 hours propyzamide treatment (T0) and 10 min after removal of propyzamide (T10) was executed. The number of identifications after filtering relative to a large background list are represented for the AP-MS results. *APC12 was retrieved with MaxQuant analysis relative to a control data set using the Araport11 database with APC12 (Q8H1U3) protein sequence from UniProt added. TAP results are derived from Van Leene et al., 2010. Putative degrons are predicted based on the APC/C degron repository (Davey and Morgan, 2016). References for ubiquitylation events are represented in the last column.

		AP-MS						TAP			Degrons			Ubiquitylation
		CCS52A1		CCS52A2		CCS52B								
Accession	Protein name	T0	T10	T0	T10	T0	T10	CCS52A1	CCS52A2	CCS52B	D-box	KEN-box	ABBA	
AT4G22910	CCS52A1	3	3					X						
AT4G11920	CCS52A2			3	3				X					
AT5G13840	CCS52B					3	3			X				
AT5G05560	APC1	3	3	3	3	3	3	X	X	X				
AT2G04660	APC2	3	3	3	3	3	3	X	X	X				
AT2G20000	APC3b/HOBBIT	3	3	3	3	3	3	X	X	X				
AT3G16320	APC3a/CDC27a	2	3	3	3	3	3							
AT4G21530	APC4	3	3	3	3	3	3	X	X	X			ABBA(1)	
AT1G06590	APC5	3	3	3	3	3	3	X	X	X				
AT1G78770	APC6/Nomega	3	3	3	3	3	3		X	X				Walton et al. 2016
AT2G39090	APC7	3	3	3	3	3	3	X	X	X				
AT3G48150	APC8	3	3	3	3	3	3	X	X	X				
AT2G18290	APC10	1	1	3		1	2		X					
Q8H1U3*	APC12	3	3	3	3	3	3							
AT1G73177	APC13	2	3	3	3	3	3							
AT5G63135	APC15	1	3	3	1	3	2							
AT2G42260	UVI4/PYM			3		3	3			X	D-box(1)			
AT3G57860	OSD1/UVI4-Like/GIGAS	3	3	3	3	3	3		X	X	D-box(1)	KEN-box(1)		
AT1G32310	SAMBA		1	3	1	3	3							
AT1G53140	DRP5A	3	3	3	3	3	3					KEN-box(1)		
AT2G45700	sterile alpha motif (SAM) domain-containing protein	2	3	3	3	3	3							
AT1G03780	TPX2	2	3	3	2	3	3				D-box(1)	KEN-box(2)		
AT1G34355	PS1	1	3	3	1	3	3					KEN-box(2)		Maor et al. 2007
AT4G14310	Transducin/WD40 repeat-like superfamily protein	1	1	3	2	3	3				D-box(2)	KEN-box(2)		Aguilar-Hernandez et al. 2017
AT2G27970	CKS2	1	2	3	2	2	3		X	X				Walton et al. 2016
AT3G48750	CDKA;1			3		3	3							Kim et al. 2013
AT5G43080	CYCA3;1			1		3					D-box(1)	KEN-box(1)	ABBA(1)	
AT1G47230	CYCA3;4			3		1					D-box(1)			

AT5G19330	ARIA	3	3	3						
AT1G22730	MRF2	2				3				
AT4G01370	MPK4			3		3	3			Kim et al. 2013
AT2G20480	unknown protein			1		3	2			
AT4G32830	AUR1			1		2		D2-box(1)		
AT4G18800	RABA1d	1	3							Kim et al. 2013
AT5G23900	Ribosomal protein L13e family protein		3							Kim et al. 2013/ Walton et al. 2016
AT5G46430	Ribosomal protein L32e		3							Kim et al. 2013/ Walton et al. 2016
AT4G05520	EHD2			3						
AT4G17650	Polyketide cyclase			3						
AT3G07090	PPPDE putative thiol peptidase family protein			2						Walton et al. 2016
AT5G48810	ATB5-B				2					
AT5G03740	Histone deacetylase 2C				3					
AT1G31860	HISN2				2					
AT5G54500	FQR1				2					
AT5G11510	MYB3R-4					3	3		KEN-box(1)	
AT1G51690	ATB ALPHA					3	3			Kim et al. 2013
AT3G58500	PP2A-4					3	2			Kim et al. 2013
AT5G60930	KIF4A ortholog					3	1	D-box(1)	KEN-box(2)	
AT4G18600	WAVE5					3	1		KEN-box(1)	
AT2G25880	AUR2					1	2	D2-box(1)		
AT3G14740	RING/FYVE/PHD zinc finger superfamily protein					1	2			
AT3G09880	ATB~ BETA					1	1			
AT5G23910	KIF2C ortholog					3			KEN-box(1)	
AT4G27060	TORTIFOLIA 1					3				
AT4G17330	G2484-1					3				Walton et al. 2016
AT5G25590	Protein of unknown function					2				
AT1G14510	AL7					2				Kim et al. 2013
AT2G24200	LAP1						2			Maor et al. 2007
AT4G24690	NBR1					3			KEN-box(1)	Kim et al. 2013 / Sarraco et al. 2009

Kim and coworkers also showed that the proteasomal inhibitor MG132 ensures stabilization of some proteins with identified ubiquitination sites (MPK4, ATB ALPHA, AL7 and NBR1), while other ubiquitinated proteins (the CCS52A1-specific interacting proteins and PP2A-4) are not stabilized by MG132 treatment (Kim et al., 2013). Looking at the expression profile of the CCS52 interacting proteins reveals that several proteins show a mitotic upregulation, similar to that of CCS52B in synchronized *Arabidopsis* suspension cultures (figure 4) (Menges and Murray, 2002). All these observations indicate the reliability of our CCS52 AP-MS data on propyzamide synchronized cell cultures.

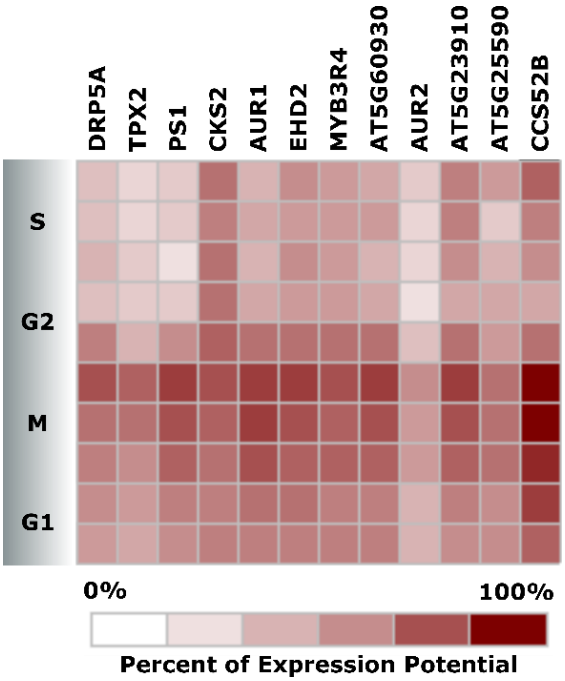


Figure 4: Heatmap of mitotic upregulated gene expression for different CCS52 interacting proteins. Expression profile based on aphidicolin synchronized Arabidopsis cell suspension cultures (Menges and Murray, 2002). Heatmap generated with Genevestigator (Hruz et al., 2008).

AP-MS on CCS52 allows pull-down of known APC substrates.

Some of the mitotic substrates identified in mammals for APC^{CDH1} (Zhou et al., 2016) are present in our data set. It has been shown that , Aurora A and Aurora B are degraded by the APC^{CDH1} during late mitosis (Stewart and Fang, 2005, 2005; Floyd et al., 2008). Here we also identify the plant Tpx2, Aurora 1 and Aurora 2. While Tpx2 is co-purified by all three CCS52 proteins, Aurora 1 is interacting with CCS52A2 and CCS52B, and Aurora 2 is specifically co-purified by CCS52B. Although Aurora 1 and 2 do not contain the classical APC degrons, a D2-type destruction box is present and has been shown in mammals to be necessary for APC^{CDH1} dependent targeting and protein degradation during anaphase (Arlot-Bonnemains et al., 2001; Littlepage and Ruderman, 2002; Demidov et al., 2005; Stewart and Fang, 2005; Floyd et al., 2008).

As mentioned before the main APC substrates are cyclins (den Elzen and Pines, 2001; Kaspar et al., 2001; Fulop et al., 2005; Wolthuis et al., 2008; Boudolf et al., 2009),

indicating that the identification of CycA3;1 and CycA3;4 in our data set is no coincidence. Although both cyclins are co-purified with CCS52A2 and CCS52B, there seems to be a preference. CycA3;1 only associates once with CCS52A2 while it is identified in all three pull-downs with CCS52B at time point zero (T0, before release from propyzamide). CycA3;4 shows the opposite association, once identified with CCS52B and three times with CCS52A2 at time point zero. The fact that both of them are not present 10 min after removal of propyzamide suggest an early and rapid degradation during mitosis, as has been reported in mammalian cells (den Elzen and Pines, 2001).

Also mitotic kinesins belong to the list of APC substrates in mammalian cells. In 2014 Singh and co-workers used multiplexed quantitative proteomics and biochemical studies to identify new APC substrates (Singh et al., 2014). Among these new candidates a group of kinesins (KIFC1, KIF18A, KIF2C, and KIF4A) was identified. In our data set two kinesins have been co-purified specifically with CCS52B, being AT5G60930 and AT5G23910. AT5G23910 is the plant ortholog of KIF2C, while AT5G60930 is the plant ortholog of KIF4A. Both kinesins are upregulated during mitosis in *Arabidopsis* with AT5G60930 showing the highest expression at M-phase of all mitotic kinesins (Vanstraelen et al., 2006). While AT5G60930 contains one putative D-boxes and two KEN-motifs, AT5G23910 only contains one putative KEN-motif.

Identification of CCS52B interactions via proximity labeling.

We also tested a proximity-based labeling method to identify the CCS52B interactome during metaphase-anaphase transition. TurboID, an altered Escherichia coli biotin ligase BirA (Branon et al., 2018; Doerr, 2018), was fused to CCS52B and transformed in our PSB-D cell suspension culture. Cells were synchronized for 4 hours with propyzamide and boosted with biotin for 2 hours (figure 5A). To make a distinction between stable interactions and substrates, the proteasome inhibitor MG132 was added simultaneously with biotin to one half of the TurboID-CCS52B cells, preventing degradation of ubiquitinated proteins (figure 5B). Combining proximity labeling and MG132 treatment was previously described by Coyaud et al. (2015) for identification of $SCF^{\beta\text{-}TrCP1/2}$ E3 ligase substrates (Coyaud et al., 2015). To allow for a label-free quantitative MS analysis three repeats with and without MG132 were conducted. All significant enriched proteins are represented in table 2 and originates from the quantitative comparison (MaxQuant) between the bait protein CCS52B and a control data set containing various in-house TurboID analyses on unrelated baits. Because TPLATE was strongly enriched in the CCS52B samples, endocytosis-related baits (TPLATE, TML and AP2M) were removed from the control data set.

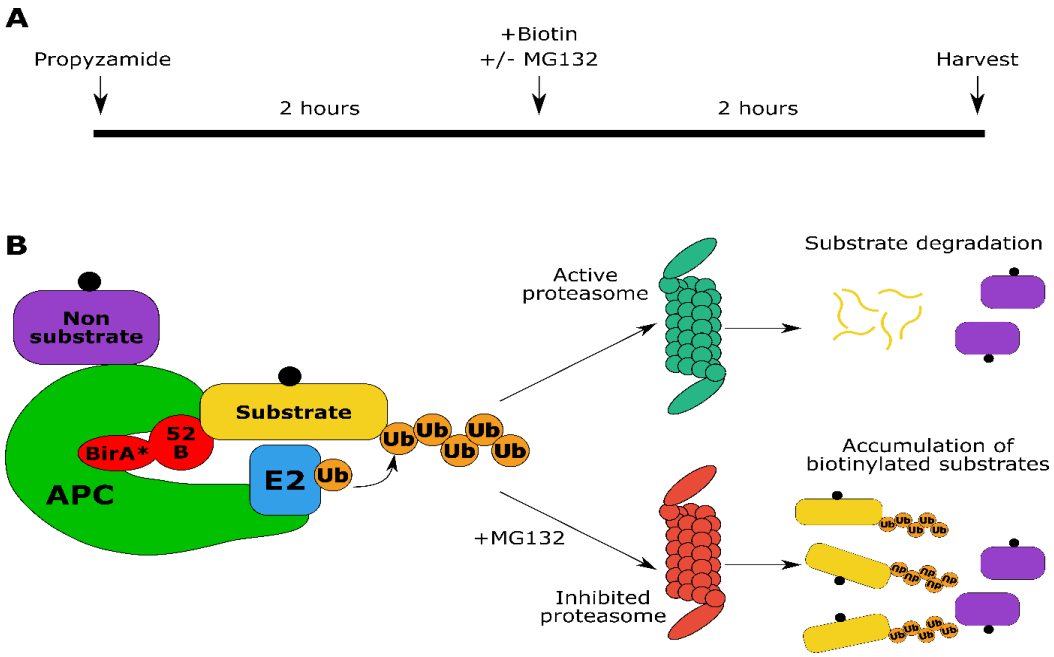


Figure 5: TurboID based proximity labeling of CCS52B interactome. **A)** Time schedule of TurboID-CCS52B transformed cell culture treated with propyzamide for 4 hours, adding the last 2 hours biotin with or without MG132. **B)** Scheme for the identification of APC substrates. Biotinylation occurs for both substrates and non-substrates. Ubiquitination of substrates results in proteasomal degradation, unless the proteasome inhibitor MG132 is added. Identification of substrates with mass spectrometry will only be possible after MG132 treatment while non-substrates will be identified under both conditions. Figure based on (Gingras et al., 2019).

Several biotinylated APC subunits and the regulator OSD1 are co-purified, however not all APC subunits are present in the final MS data set. Probably because these subunits are hidden in the intricate macromolecular APC complex and thus not accessible for biotinylation. Next to the APC subunits and regulator, only one protein is co-purified with both AP-MS and TurboID, which is PS1. Identification of PS1 with and without MG132 treatment suggests that this protein has a regulatory function towards APC. Other MG132 independent CCS52B interactors are TPLATE, a cytokinesis protein targeted to the cell plate (Van Damme et al., 2004), and some mitotic unrelated or unknown proteins.

Our data also confirms the interaction of CCS52B with CycB1;1 (Fulop et al., 2005). However, with MG132 treatment, CycB1;1 is not co-purified, which is contradicting the fact that CycB1;1 is a substrate of the APC (Kwee and Sundaresan, 2003; Rojas et al., 2009; Eloy et al., 2011). Also no other cyclins are biotinylated, which is not what we expected.

Table 4: Proteins identified by quantitative liquid chromatography-tandem MS (LC-MS/MS) after TurboID proximity labeling on the APC co-activator CCS52B. Three TurboID analyzes with and without MG132 were executed. A comparative analysis was performed against a TurboID control data set in Perseus using the MaxQuant LFQ intensity values. Fold change (Difference) and significance (-log(p-value)) are represented for significantly enriched proteins with a false discovery rate of 0.05 and a S0 of 0.5. * APC12 was retrieved with MaxQuant analysis relative to the TurboID control data set using the Araport11 database with APC12 (Q8H1U3) protein sequence from UniProt added. Putative degrons are predicted based on the APC/C degron repository (Davey and Morgan, 2016). References for ubiquitylation events are represented in the last column.

		TurboID							
		-MG132		+MG132		Degrons			Ubiquitylation
Accession	Protein name	Difference	-log(p-value)	Difference	-log(p-value)	D-box	KEN-box	ABBA	
AT5G13840	CCS52B	16.80	26.16	16.32	25.48				
AT5G05560	APC1	6.93	21.65	4.06	11.77				
AT3G16320	APC3a/CDC27a	6.57	18.94	5.51	15.53				
AT2G20000	APC3b/HOBBIT	9.86	17.75	9.15	16.25				
AT1G06590	APC5	5.37	12.51	4.92	11.47				
AT2G39090	APC7	10.57	28.33	9.55	25.85				
AT3G48150	APC8	4.23	5.60						
Q8H1U3*	APC12	*	*	*	*				
AT3G57860	OSD1/UVI4-Like/GIGAS	7.32	23.15	6.22	19.39				
AT4G37490	CYCLIN B1;1	3.36	9.68			D-box(4)			
AT1G34355	PS1	3.07	7.96	7.80	25.49	D-box(2)	KEN-box(2)		Maor et al. 2007
AT3G46540	ENTH/VHS family protein	6.20	11.60	7.48	15.01		KEN-box(1)		
AT3G01780	TPLATE	11.48	21.28	7.32	12.44	D-box (2)			Kim et al. 2013
AT3G27400	PLL18	4.76	15.02	4.89	16.37				
AT2G42230	C-CAP/cofactor C-like domain-containing protein	3.28	7.59	3.77	9.27	D-box(1)	KEN-box(1)		
AT1G25260	Ribosomal protein L10 family protein	2.90	9.97	3.44	12.59				
AT1G79880	La2	3.92	8.97	3.04	6.14		KEN-box(1)		
AT5G52560	USP	2.94	8.33	2.89	8.01				
AT1G36180	ACC2	3.46	6.59			D-box(2)			Kim et al. 2013
AT5G09330	NAC082	2.58	8.93						
AT5G57130	SMXL5	2.71	9.57			D-box(1)			

AT3G47890	Ubiquitin carboxyl-terminal hydrolase	3.25	5.90			D-box(5)		Maor et al. 2007
AT3G13460	ECT2			4.32	13.89			
AT3G42660	EOL1			4.22	7.00			
AT2G28510	DOF2.1			4.07	16.13			
AT3G25150	Nuclear transport factor 2 (NTF2) family protein			3.72	5.75			Walton et al. 2016
AT1G71730	hypothetical protein			3.55	8.17			
AT1G79000	HAC1			3.55	7.92		KEN-box(1)	
AT3G17750	DYRKP-1			3.52	4.74	D-box(1)		
AT1G67250	Proteasome maturation factor UMP1			3.48	12.99			
AT3G13990	Dentin sialophosphoprotein			3.48	4.65			Kim et al. 2013 / Walton et al. 2016 / Aguilar-Hernandez et al. 2017
AT3G52560	UEV1D-4			3.39	5.34			
AT5G49555	FAD/NAD(P)-binding oxidoreductase family protein			3.37	4.95	D-box(2)		
AT5G26800	xaa-pro aminopeptidase P			3.33	7.91	D-box(1)		
AT3G52250	PWR			3.22	7.04			
AT3G23830	RBGA4			2.64	6.96			
AT5G49160	MET1			2.46	7.67			
AT1G43140	Cullin family protein			2.44	7.75	D-box(1)		
AT2G28760	UXS6			4.70	18.98			Maor et al. 2007/ Kim et al. 2013 / Walton et al. 2016

Some of the biotinylated proteins contain putative major degrons (table 2), and not many are linked with cell division and cytokinesis. Next to CycB1;1, PS1 and TPLATE, EOL1 is expressed in dividing cells and is necessary for the maintenance of H3K27me3 modifications and epigenetic memory of gene repression (Zhou et al., 2017). Another biotinylated protein involved in cell division is DOF2.1, a transcription factor specifically controlling procambium cell divisions in roots (Smet et al., 2019). EOL1 and DOF2.1 are only identified after blocking the proteasomal degradation by MG132, but the absence of major degrons leaves it doubtful if EOL1 and DOF2.1 are substrates of APC^{CCS52B}. Proteome-wide ubiquitination studies (Maor et al., 2007; Saracco et al., 2009; Kim et al., 2013; Walton et al., 2016; Aguilar-Hernandez et al., 2017; Willems et al., 2019) reveal ubiquitination sites on seven biotinylated proteins, among which PS1 and TPLATE (table 2).

Discussion

The anaphase-promoting complex is the most intricate E3 type mainly functioning during cell division. This E3 ubiquitin-protein ligase relies on its co-activators, CDC20 and CCS52, for ubiquitination of specific substrates in a time-dependent manner. While the list of APC targets in mammals and yeast is comprehensive (Kramer et al., 2000; Zhou et al., 2016), the amount of known substrates for the plant CDH1 ortholog CCS52 is limited (Heyman and De Veylder, 2012). Here we performed AP-MS on *Arabidopsis* cell suspension culture for all three CCS52 co-activators to further supplement the list of APC^{CCS52} substrates during cell division. To enrich for an active APC^{CCS52} complex during mitosis, and more specific during metaphase-anaphase transition, we treated cells with propyzamide. This anti-tubulin drug allowed for metaphase arrest in approximately ¼ of the *Arabidopsis* cells and a synchronized continuation of mitosis after removal. AP-MS on GS^{rhino}-CCS52 synchronized cell cultures revealed common and specific interacting proteins for the three CCS52 co-activators.

AP-MS is more efficient than TAP-MS and allows pull-down of known E3 ligase substrates.

Previous tandem affinity purifications to unravel the core cell cycle machinery has allowed for the pull-down of all three CCS52 co-activators together with subunits of the APC complex and even a CCS52A2 specific substrate (Van Leene et al., 2010; Heyman et al., 2013). However, performing TAP-MS did not allow the identification of the complete APC complex, partly due to the less sensitive mass spectrometry at that time. Our AP-MS method however, allowed for the identification of all subunits but one, APC11, and known regulators (UVI4, OSD1 and SAMBA). Even the smallest subunits APC13 (7,14 kDa) and APC15 (11,23 kDa) were identified, as well as the recently characterized APC12 (cdc26, 7,22 kDa) subunit (Lorenzo-Orts et al., 2019). The importance of APC13 and APC15 for efficient cyclin degradation during anaphase has been demonstrated in plants and human cells (Schwickart et al., 2004; Mansfeld et al., 2011). Our AP-MS data also contain known substrates and a lot of proteins with putative degrons which were not co-purified with TAP-MS. The presence of known APC^{CDH1} substrates proves that with our AP-MS it is possible to pick up transiently interacting substrates for an E3-ligase.

The main APC substrates are cyclins (den Elzen and Pines, 2001; Kaspar et al., 2001; Fulop et al., 2005; Wolthuis et al., 2008; Boudolf et al., 2009), and also in our data set cyclins are identified, namely CycA3;1 and CycA3;4. Interaction of CycA3;4 with the CCS52 co-activators has been demonstrated previously with a Y2H analysis

demonstrating a strong interaction with CCS52A1 and CCS52B, and a weak interaction with CCS52A2 (Fulop et al., 2005). However, our data suggest a preference of CCS52A2 for CycA3;4. Next to our data, CCS52A2 dependent degradation of CycA3;4 in *Arabidopsis* has recently been postulated by Willems A. (unpublished data). *Arabidopsis* plants expressing CycA3;4-GUS under the endogenous promoter in a ccs52a2 mutant background show enhanced stabilization of CycA3;4. Furthermore, MG132-treated CycA3;4-GUS lines show a strong increase in GUS levels. Also CycA3;4 overexpressing lines, that have a dwarfed phenotype, are more or less rescued by crossing with a CCS52A2 overexpressing line. Taken ours and Willems A. *in vivo* analyses together shows that analyzing binary interactions, like Y2H screens, do not always give information about the function of the interaction. The strong Y2H interactions can suggest a regulatory function for CycA3;4 relative to CCS52A1 and CCS52B, while association with CCS52A2 results in ubiquitination and 26S proteasomal degradation of CycA3;4. While CycA3;4 is preferentially bound by CCS52A2 in our AP-MS data set, CycA3;1 interaction is mainly identified with CCS52B. APC dependent CycA3;1 degradation has previously been demonstrated in apc2 mutant Arabidopsis embryo's and CCS52A1-silenced BY2 protoplast showing enhanced CycA3;1 stabilization (Capron et al., 2003; Mathieu-Rivet et al., 2010). Both cyclins are not identified after removal of propyzamide, suggesting an early and rapid degradation. Early mitotic degradation of CycA3;4 has been observed in BY2 cells, where CycA3;4 co-localizes with condensing prophase chromatin but not with the aligned metaphase chromosomes (Boruc et al., 2010). Also CycA3;1-GFP signals disappear in BY2 cells during metaphase-anaphase transition (Boruc et al., 2010). Further in vivo analysis, similar to that of Willems A. should confirm an APC^{CCS52B} dependent ubiquitination of CycA3;1 and proteasomal degradation during mitosis.

Another identified interacting protein that has been linked with the APC and cyclins is the mitotic upregulated CKS2. CKS2 is a positive regulator of the CDK-cyclin complexes and thus essential for cell cycle control. In mammalian cells it has been demonstrated that CKS2 is essential for the first metaphase/anaphase transition of mammalian meiosis (Spruck et al., 2003; Martinsson-Ahlzen et al., 2008). TAP-MS and BiFC analyses have revealed a strong interaction of CKS2 with different CDKs and cyclins including CycA3;1, CycA3;4 and CDKA;1, present in our data set (Van Leene et al., 2007; Boruc et al., 2010). A Cyclin-CDK-CKS complex is essential for the activation of the APC complex during onset of mitosis by phosphorylation of APC3 and APC1, allowing CDC20 association (Patra and Dunphy, 1998; Zhang et al., 2016; Kataria and Yamano, 2019; Yamano, 2019). As a consequence, we cannot conclude from our AP-MS data set that we only co-purified CycA3;4 and CycA3;1 as substrates, but they might be also pulled-down as regulators of the APC. Furthermore, a cyclin–

CDK–CKS complex is also needed for efficient APC dependent degradation of cyclins in mammalian cells (Wolthuis et al., 2008; van Zon et al., 2010). Prior to degradation, cyclin A and cyclin B1 are directed to the phosphorylated APC^{CDC20} by its CDK and CKS partners during prometaphase, independent of CDC20 or D-box motifs. This recruitment to the APC is CKS and APC3 dependent. Without CKS, degradation of the cyclins is delayed (van Zon et al., 2010). A similar mechanism could exist for plant cyclins and APC^{CCS52B}.

We also identified Tpx2, Aurora 1 and Aurora 2, mitotic APC^{CDH1} substrates identified in mammals (Stewart and Fang, 2005, 2005; Floyd et al., 2008). TPX2, a microtubule associated protein, is a key regulator of mitotic spindle assembly during late prophase and early prometaphase by binding and activating Aurora A (Vos et al., 2008; Garrido and Vernos, 2016). TPX2 specific degradation by APC^{CDH1} at anaphase and during cytokinesis in human cells has been demonstrated *in vitro* as well as *in vivo* (Stewart and Fang, 2005). Aurora 1 interacts with CCS52A2 and CCS52B, and Aurora 2 is specifically co-purified by CCS52B. Although Aurora 1 and 2 do not contain the classical APC degrons, a D2-type destruction box is present and has been shown in mammals to be necessary for APC^{CDH1} dependent targeting and protein degradation during anaphase (Arlot-Bonnemains et al., 2001; Littlepage and Ruderman, 2002; Demidov et al., 2005; Stewart and Fang, 2005; Floyd et al., 2008).

Other known APC substrates in our AP-MS data set are KIF2C (AT5G23910) and KIF4A (AT5G60930). KIF2C is not a typical kinesin that carries cargo along the microtubule. Instead, it localizes at the centromeres, kinetochores and spindle poles during early mitosis and depolymerizes microtubules with ATPase activity. This is essential for correct chromosome movement and segregation during metaphase. Depletion of KIF2C in HeLa cells resulted in a block of the prometaphase to metaphase transition with misaligned chromosomes (Zhu et al., 2005). Furthermore, KIF2C disappears from the spindle poles at metaphase-anaphase transition (Ganguly et al., 2008) indicating an APC dependent degradation, confirmed by Singh et al. (2014). Singh and co-workers used multiplexed quantitative proteomics and biochemical studies to identify new APC substrates (Singh et al., 2014). Among these new candidates a group of kinesins (KIFC1, KIF18A, KIF2C, and KIF4A) was identified. Also KIF4A plays a role during chromosome segregation and is necessary for a normal metaphase chromosome morphology in mammalian cells (Mazumdar et al., 2004). Localization studies in mammalian cells revealed a clear association with the chromosomes during the entire M-phase but also with the central spindle midzone that is formed during late anaphase (Kurasawa et al., 2004; Zhu and Jiang, 2005). This central spindle midzone is essential for a correct cytokinesis and functions as a binding

site for different proteins that are responsible for cell cleavage. Formation of this central spindle midzone is dependent on KIF4A and its binding partner PRC1 (the AtMAP65 ortholog) (Zhu and Jiang, 2005). Also in *Arabidopsis* a compaction of AtMAP65 in the spindle midzone is observed, suggesting that the KIF4A ortholog AT5G60930 has a similar function as in mammalian cells (Mao et al., 2005). Further functional characterization of these mitotic kinesins should reveal if they have a similar function as in mammalian cells and their specific ubiquitination by the APC.

Common CCS52 interacting proteins act as regulators and/or substrates.

Almost all AP-MS identified interacting proteins have a link with mitosis, cytokinesis or associate with microtubuli. Proteins which are co-purified with all three CCS52 proteins are TPX2, CKS2, DRP5A, PS1, AT2G45700 and AT4G14310. As already mentioned before, TPX2 and CKS2 have a clear link with the APC. DRP5A is a dynamin protein expressed only in dividing cells at M-phase and localizes at the forming cell plate during cytokinesis (Menges and Murray, 2002; Miyagishima et al., 2008). DRP5A does contain a putative KEN-box and could be a direct target of the APC. APC-dependent degradation of dynamins has been observed in mammalian cells. DRP1, a dynamin protein involved in mitochondrial fission during cell division, was shown to be degraded by APC^{CDH1} in a D-box dependent way (Horn et al., 2011).

AT2G45700 encodes for a protein that shows homology at the C-terminal part for the mammalian ortholog SNM1. SNM1 is required for cell cycle arrest during early prophase in response to aberrant mitotic spindle formation (Akhter et al., 2004). It strongly and constitutively binds with the APC complex, functioning as negative regulator of APC during the early mitotic checkpoint in mammalian cells. Identification of AT2G45700 with all three CCS52 proteins in our cell culture upon exposure to the spindle poison propyzamide, suggest a similar function in *Arabidopsis*.

AT4G14310 is an unknown plant specific protein, containing putative D-box and KEN-box motifs. Interaction studies have shown that this unknown protein associates with the core cell cycle machinery (Van Leene et al., 2010), strongly interacting with CDKA, CycA3;1, CycA3;4 and CKS2, all present in our AP-MS data set. TAP performed on AT4G14310 also revealed interaction with the division plane marker KCA2 (Van Leene et al., 2010). Subcellular localization of AT4G14310-GFP in tobacco BY-2 cells revealed its association with mitotic microtubules until anaphase/telophase transition (supplementary figure 2, unpublished data from Gadeyne A.). AT4G14310 localizes to mitotic structures present in the division plane like the PPB, the mitotic spindle and the early phragmoplast. A hypothesis is that AT4G14310 recruits cell cycle proteins (like

CDKA;1) to the division zone resulting in the regulation of division zone markers (like KCA2). In addition, its cell cycle dependent degradation seems coordinated with the inactivation of CDKA;1 at anaphase/telophase transition (unpublished data from Gadeyne A.), suggesting a role for the APC.

PS1 has been linked with mitotic spindle orientation. The mutant version shows abnormal spindle orientation at male meiosis II leading to the production of diploid pollen grains (d'Erfurth et al., 2008). The expression profile of PS1 is similar to that of CCS52B and CDC20.2 with a peak at M-phase (figure 3) (Menges et al., 2005). The presence of two KEN-boxes and ubiquitination sites (Maor et al., 2007) suggests an APC dependent degradation. In addition, performing proximity labeling on CCS52B resulted in the identification of PS1 with and without the proteasome inhibitor MG132, showing a stabilization of PS1 under MG132 treatment. Further characterization of PS1 and its different destruction motifs is needed to validate PS1 as an APC substrate.

A mitotic function for CCS52B.

Based on expression profiles and localization studies it is postulated that CCS52B plays a main role during mitosis and thus fulfills the role of CDH1 in *Arabidopsis* (Menges et al., 2003, Yang et al., 2017). Our data further confirms this hypothesis. From the 57 proteins co-purified with all three co-activators 77% interacts with CCS52B. Fifteen proteins specifically interact with CCS52B, and all but one contain putative major degrons. Most of them are linked with mitosis, and are thus putative substrates of APC^{CCS52B}. However, we must keep in mind that overexpression of the co-activators can affect the strong cell cycle related regulation and may result in the overrepresentation of putative interactors. Similarly, propyzamide synchronization leads to an enrichment of mitotic proteins and may also result in overrepresentation of putative CCS52 interactions. Therefore, validation is crucial to state that AP-MS on the CCS52 co-activators has resulted in the identification of new mitotic substrates. However, the presence of known APC^{CDH1} substrates proves that with our AP-MS technique it is possible to pick up transiently interacting substrates for an E3-ligase.

We previously discussed the two mitotic kinesins (AT5G60930 and AT5G23910) and Aurora 2 as putative APC^{CCS52B} substrates based on their orthologs in mammalian cells (Demidov et al., 2005; Kawabe et al., 2005; Singh et al., 2014). AP-MS on the CCS52B co-activator resulted also in the specific isolation of subunits of a protein phosphatase type 2A (PP2A) complex, PP2A-4, ATB ALPHA and ATB' BETA. PP2A complexes are major phosphatases involved in nearly every cellular process (Seshacharyulu et al., 2013; Wlodarchak and Xing, 2016; Kataria and Yamano, 2019; Mathe et al., 2019).

They comprise a catalytic C subunit (e.g. PP2A-4), a scaffolding A subunit and a regulatory B subunit (e.g. ATB' BETA and ATB ALPHA), which controls substrate selectivity and subcellular localization of the enzyme. A link between different PP2As and mitosis has been established in plants as well as in mammalian cells. PP2As have been linked with the formation of the preprophase band (PPB) in *Arabidopsis* (Spinner et al., 2013), and in mammalian cells a PP2A complex helps fine-tuning the metaphase-to-anaphase transition by tightly controlling separase activation (Hellmuth et al., 2014). Also mitotic dephosphorylation of CDC20 by a PP2A has been demonstrated to activate APC at early mitosis (Lee et al., 2017). The ATB ALPHA homologs in human cells, the B55 family, are responsible for the dephosphorylation of CDH1 at anaphase, stimulating co-activator exchange of CDC20 to CDH1 (Kataria and Yamano, 2019). Furthermore, the yeast homolog of ATB ALPHA, CDC55, dephosphorylates APC^{CDC20} during the spindle assembly checkpoint, resulting in inactivation of the APC as long as sister kinetochores are not properly attached by spindle microtubules and aligned at the metaphase plate (Rossio et al., 2013). Further characterization of this CCS52B specific PP2A complex should shed light on its function during mitosis and if it regulates the phosphorylation status of the APC or the co-activators.

The MYB3R-4 transcription factor is another interactor of CCS52B. This TF activates the expression of G2/M-phase specific genes (like CycA and CycB) under normal growth conditions (Haga et al., 2007). Although under DNA stress MYB3R-4-GFP levels rapidly decrease in *Arabidopsis* roots, MG132 treatment did not change the protein level of MYB3R4-GFP under normal growth conditions, suggesting degradation of MYB3R-4 is not dependent on the 26S proteasome (Chen et al., 2017). However, one should think that early mitotic degradation of MYB3R-4 is needed to prevent expression of cyclins at later stages of cell division. Our AP-MS data and the presence of a putative degron, postulates that MYB3R-4 is targeted by APC^{CCS52B} at metaphase-anaphase transition. Therefore, further characterization of MYB3R-4 and its protein levels during the different mitotic phases is needed.

Another CCS52B, and CCS52A2, interactor is MPK4, a kinase that is recruited at the forming cell plate and is essential for progression of cytokinesis by phosphorylating and deactivating the MICROTUBULE ASSOCIATED PROTEIN (MAP) 65-3 (Kosetsu et al., 2010; Buschmann and Muller, 2019). Mutant *mpk4* plants have a prolonged metaphase, anaphase and cytokinesis with incomplete phragmoplast expansion (Beck et al., 2011). Ubiquitination of MPK4 has been observed in *Arabidopsis* (Kim et al., 2013). Moreover, MG132 treatment of *Arabidopsis* seedling resulted in a stabilization of MPK4 ubiquitination, meaning that MPK4 undergoes 26S proteasomal degradation.

Future *in vitro* and *in vivo* studies will reveal if the APC is responsible for MPK4 ubiquitination and if this is specifically performed by CCS52A2 or CCS52B.

NBR1 interacts with CCS52B during metaphase arrest. NBR1 is able to bind ubiquitin and probably has a function in autophagosomal degradation of ubiquitinated proteins (Svenning et al., 2011). However, biochemical studies have revealed that NBR1 has functional characteristics similar to mammalian p62 (Svenning et al., 2011), a regulator of cyclin B1 levels and crucial for a proper mitosis (Linares et al., 2011). Ubiquitination of NBR1 has been confirmed by TAP-MS in Arabidopsis and is stabilized by the proteasomal inhibitor MG132 (Saracco et al., 2009; Kim et al., 2013). Further in vivo analysis is needed to proof its mitotic function in Arabidopsis and its APC dependent ubiquitination and degradation.

Other CCS52B specific interactors are WAVE5, a subunit of the WAVE complex that is responsible for actin cytoskeleton organization (Brembu et al., 2004), and some other less characterized or unknown proteins like AT2G20480, a plant-specific peptide which also has been co-purified after pull-down of SAMBA in maize (unpublished data, Bontinck M.).

Specific CCS52A-type interacting proteins have a link with mitosis.

The amount of CCS52A1 and CCS52A2 specific interacting proteins after AP-MS is very limited compared to the list of specific CCS52B interacting proteins. However, two of them have a link with mitosis or cytokinesis. The CCS52A1 specific interactor RabA1d is a small GTPase involved in cell plate formation (Berson et al., 2014). Localization studies in root cells revealed that RabA1d accumulates at the nascent cell plate and later on follows the expanding cell plate in a disc-like structure. Ubiquitination of RabA1d has been observed but the protein level is not affected by MG132 treatment, and therefore not subjected for proteasomal degradation. In addition, RabA1d does not contain major degrons suggesting that its interaction with $APC^{CCS52A1}$ is indirect and/or RabA1d has a regulatory function towards CCS52A1.

EHD2 was specifically co-purified with CCS52A2 and has been linked with endocytosis (Bar et al., 2008). The mammalian ortholog, EHD1, has in addition to its function in endocytosis, also a role in controlling central spindle formation at anaphase and recruiting endosomes to the intercellular bridge during cytokinesis (Reinecke et al., 2015). Further *in vitro* and *in vivo* analysis is needed to proof a mitotic function of EHD2 in *Arabidopsis* and its APC dependent ubiquitination and degradation.

Validation of putative APC^{CCS52} substrates.

AP-MS of the APC co-activators during metaphase/anaphase transition has resulted in the identification of several candidate substrates. Some of them have been confirmed as substrates in other species or have been linked to mitosis. But further validation is needed if we want to prove the applicability of our AP-MS protocol. Especially the validation of CCS52B targets is of great interest, because little is known about the biological relevance of CCS52B (Heyman and De Veylder, 2012). To demonstrate that AP-MS allows for identification of APC substrates we are setting up *in vitro* ubiquitination assays and *in vivo* stabilization studies on specific CCS52 interacting proteins. We depicted the CCS52A1 specific interactor RabA1d, the CCS52A2 specific interactor EHD2 and the CCS52B specific interactors AUR2, NBR1 and MPK4.

Recombinant proteins fused to Maltose-binding protein (MBP) have been purified from *E. coli*, except for NBR1. In addition, purification has been obtained for two recombinant E2 ubiquitin-conjugating enzymes, UBC19 and UBC20, which have been proposed to specifically operate with the APC complex (Criqui et al., 2002). We tested an *in vitro* ubiquitination assay, combining human E1, recombinant E2 and GS^{rhino}-CCS52 TAP or AP purified APC complex with the recombinant putative substrates. However, we could not detect any ubiquitination events, not even for the known substrate AUR2. Although similar studies have been successfully performed with purified APC complex from human cells (Amador et al., 2007; Garnett et al., 2009; Song et al., 2011), we doubt the activity of our APC purified complex. Looking back at our AP-MS data, we retrieve all APC subunit, except for APC11 which is one of the two core subunits for catalytic activity. Together with the other catalytic core subunit APC2, they are sufficient to catalyze non-specific ubiquitination reactions *in vitro* (Gmachl et al., 2000; Tang et al., 2001). The absence of APC11 after AP-MS or TAP (Van Leene et al., 2010), can explain our negative *in vitro* ubiquitination results. Furthermore, we also detect a relative high amount of the negative APC regulator OSD1 (normalized spectral abundance factor (NSAF) of around 0.3) after AP-MS, which will also interfere with the APC activity during the *in vitro* ubiquitination assays. At the moment, we are testing an alternative *in vitro* ubiquitination assay using protein extracts of wild type and mutant ccs52 *Arabidopsis* lines, based on previous described cell free degradation assays (Wang et al., 2009; Lin et al., 2012; Lin et al., 2015).

To further validate the CCS52 specific interacting proteins *in vivo*, protein stabilization will be analyzed in all three ccs52 mutant *Arabidopsis* backgrounds. Transgenic GFP-RabA1d (Berson et al., 2014) and MPK4-YFP (Kosetsu et al., 2010) plants have been crossed with homozygous ccs52 mutant lines. Constructs with RabA1d, EHD2, AUR2

and NBR1 fused to GFP and under the CDKA;1 promoter have been transformed in wild type and ccs52 mutant backgrounds. In the near future stabilization of the putative APC targets will be assessed via GFP/YFP pull-down and subsequent protein abundancy visualization on western blot. In parallel, confocal microscopy will be performed to reveal differences in localization and protein abundancy in the different mutant backgrounds.

Proximity labeling does not validate AP-MS.

We were unsuccessful to confirm the CCS52B AP-MS interactome by proximity labeling. Despite the absence of different cyclins, interesting hits were identified like PS1, CycB1;1 and TPLATE. As mentioned before, we postulate that PS1 has a regulatory role with respect to the APC. Our data also suggests a stable interaction of TPLATE with CCS52B because TPLATE is biotinylated in a MG132 independent way. TPLATE, a cytokinesis protein targeted to the cell plate and involved in cell plate anchoring (Van Damme et al., 2004; Van Damme et al., 2006), could also be biotinylated because CCS52B protein levels peak during cytokinesis (Yang et al., 2017). Even so, localization studies have shown that TPLATE disappears from the post-cytokinetic wall (Van Damme et al., 2004). This in combination with the presence of two D-box motifs and observed ubiquitination events (Kim et al., 2013) suggests a APC dependent degradation of TPLATE after cytokinesis. That TPLATE degradation is dependent on the 26 proteasome has been confirmed by a MG132 stabilization study (Kim et al., 2013). Further analysis is needed to reveal a cytokinetic interaction of TPLATE with APC^{CCS52B} and/or its post-cytokinetic ubiquitination. On the other hand, interaction of CycB1;1 with CCS52B has been confirmed by a Y2H analysis (Fulop et al., 2005). Although cyclins are known substrates of the APC, our TurboID analysis does not suggest a APC^{CCS52B} dependent degradation of CycB1;1 because it was not enriched in the MG132-treated samples. It could however play a regulatory role, phosphorylating APC^{CCS52B} in collaboration with a CDK and CKS. Our data and interaction studies, showing interaction of CycB1;1 with CCS52A1, CCS52A2 and not with CDC20 (Fulop et al., 2005; Kevei et al., 2011), suggest that degradation of CycB1;1 during anaphase (Boruc et al., 2010) could be dependent of $APC^{CCS52A1/2}$.

The absence of known APC substrates with proximity labeling suggest that our setup is not optimal. However, proximity labeling experiments have been successfully performed for the identification of E3-ligase substrates (Coyaud et al., 2015; Iconomou and Saunders, 2016). The only difference is that we synchronize with propyzamide to enrich for cells in mitosis. Propyzamide blocks cells during metaphase by inhibiting the formation of the mitotic spindle. This triggers the mitotic checkpoint complex (MCC)

which blocks the APC, preventing degradation of cyclins and securin and thus the onset of anaphase (Liu and Zhang, 2016). Only when sister kinetochores are properly attached by spindle microtubules and aligned at the metaphase plate, the APC is unlocked. Here we performed proximity labeling at a time point that the APC is inhibited. More relevant data could be obtained when biotinylation is boosted after removal of propyzamide and when the APC is reactivated.

With AP-MS we identified different known and putative APC^{CCS52} substrates, although we applied the same conditions as for proximity labeling. Even at time point zero (when propyzamide blocks the formation of the mitotic spindle and APC is inhibited) we identified cyclins, Aurora kinases and mitotic kinesins. It is possible that finding these known substrates at a time point that APC is blocked *in vivo* by MCC, is the result of *in vitro* interactions occurring during protein extraction or affinity purification. By applying propyzamide, we enrich for mitotic proteins and increased the chance to identify APC^{CCS52} substrates.

However, a more plausible explanation for the suboptimal results with TurboID is the steric hindrance the large TurboID tag (+/- 46 kDa) causes. While subunits of the APC platform and structural module were biotinylated, the catalytic core subunits were not retrieved. This could suggest that TurboID-CCS52B association with the APC obstructs the formation of an active E3-ligase complex. On the other hand, it could also be that the catalytic subunits are present but are not subjected to biotinylation. It has been shown that mutating one of two essential binding domains of the co-activators, being the N-terminal C-box and C-terminal IR-tail, results in a dramatic loss of APC activity (Matyskiela and Morgan, 2009). In our case, the N-terminal fusion of TurboID on CCS52B, close to the C-box, could interfere with the proper binding of the C-box at the APC8 binding site (Yamano, 2019), and could thus diminish the APC activity. As an alternative, we could fuse TurboID at the C-terminal region of CCS52B, keeping in mind that this also could interfere with proper association of the co-activator with the APC, since the essential C-terminal IR-tail could be shielded. To prevent interference with the APC and substrate binding, it would be more ideal to position TurboID in a loop region relatively far away from the C-box, IR-tail and WD40 domain.

Future perspectives

Further validation of the putative new substrates of APC^{CCS52} is needed. *In vitro* ubiquitination analyses and *in vivo* stabilization studies in mutant ccs52 *Arabidopsis* plants have been started up for the CCS52A1 specific interactor RabA1d, the CCS52A2 specific interactor EHD2 and the CCS52B specific interactors AUR2, NBR1

and MPK4. Furthermore, characterization of the putative degrons by mutation analysis should be examined. Alongside, performing AP-MS on all five co-activators of the APC at different time points during cell cycle will allow us to further expand the APC interactome. Since CCS52B protein levels peak during cytokinesis (Yang et al., 2017), further characterization of the not so well-known CCS52B and its specific interacting proteins and substrates during later phases of mitosis is needed. In addition, TurboID-based proximity labeling experiment need to be further explored to obtain a comprehensive APC interactome. Combining the APC interactome with a cell cycle based ubiquitination profile will bring more clarity on the APC substrates. For the latter, one can use an anti-diglycyl lysine antibody to enrich for trypsin digested ubiquitinated peptides of wild type and ccs52 mutant plant protein extracts (Xu and Jaffrey, 2013).

Materials and methods

Cloning

Generation of expression constructs was obtained by following a gene stacking approach with the use of the MultiSite Gateway technology (Karimi et al., 2002). The expression vector consisting of the microtubule-binding domain (MBD) of MICROTUBULE-ASSOCIATED PROTEIN 4 (MAP4) fused to GFP and driven by the constitutive 35S promoter (Marc et al., 1998) was kindly provided by D. Van Damme. Generation of CCS52A1/A2/B and GS^{rhino} entry vectors is described elsewhere (Van Leene et al., 2007; Van Leene et al., 2015). An entry vector containing 3xHA-TurboID preceded by an omegaleader and followed by an 13xG4Slinker was obtained by DNA synthesis in the pUC57 vector (GenScript). The pK7m24GW2 destination vector can be retrieved at https://gateway.psb.ugent.be/. Subsequent MultiSite Gateway recombination generated the expression vectors pK7:35S-GS^{rhino}-CCS52A1, pK7:35S-GS^{rhino}-CCS52A2, pK7:35S-GS^{rhino}-CCS52B and pK7:35S-TurboID-CCS52B.

Cell culture transformation and propyzamide treatment

Arabidopsis cell cultures were transformed, maintained, upscaled and harvested as previously described (Van Leene et al., 2015). One day after subculturing, cell cultures were blocked at metaphase by addition of 6 µM propyzamide for 4 or 6 hours. Propyzamide block was reversed by gently removing the medium by vacuum filtration, without allowing the cells to dry. Cells were washed three times with MSMO medium and resolved in MSMO medium for further growth. To follow propyzamide synchronization, transgenic GFP-MAP4 cells were analyzed under the Axiovert inverted microscope (Zeiss). Samples were taken after 4 (or 6) hours propyzamide treatment and every 5 minutes after removal of propyzamide. In every sample 15 cell colonies of approximately 15 cells were checked for mitotic spindle and phragmoplast formation as quickly as possible. Live imaging of transgenic GFP-MAP4 cells was obtained with a Zeiss 710 inverted confocal microscope with the FV10-ASW Viewer software package and equipped with a 60x (NA. 1.2) water corrected objective. GFP was visualized using 488 nm laser excitation and 500-530 nm spectral detection and images were taken for a time period of 100 minutes

Coupling rabbit IgG to magnetic beads

Rabbit IgG antibodies (Sigma) were coupled in-house on BcMag™ Epoxy-activated Magnetic beads (Bioclone) as previously described (Hamperl et al., 2014).

Affinity purification (AP)

Triplicate affinity purification experiments followed by on-bead trypsin digestion were performed as previously described (Van Leene et al., 2019) and goes as follows. Triplicate pull-down experiments were performed on extracts derived from propyzamide synchronized GS^{rhino}-CCS52A1, GS^{rhino}-CCS52A2, GS^{rhino}-CCS52B cultures following the standard TAP extraction cell culture protocol (Van Leene et al., 2015), however for the benzonase treatment, the incubation time was shortened to 15 min. To isolate protein complexes, 25 mg total protein extract was incubated 45 min with 50 μL magnetic IgG bead suspension. Beads were washed three times with 500 μL TAP extraction buffer and one time with 500 μL TAP extraction buffer without detergent. The proteins were digested on-bead as follows. After the final wash with extraction buffer without detergent, the beads were washed with 500 μL 50 mM NH_4HCO_3 (pH 8.0). The wash buffer was removed and 50 μL 50 mM NH_4OH was added together with 1 μg Trypsin/Lys-C and incubated at 37°C for 4 h in a thermomixer at 800 rpm. Next, the digest was separated from the beads and overnight incubated with 0.5 μg Trypsin/Lys-C at 37°C. Finally, the digest was centrifuged at 20800 rcf in an Eppendorf centrifuge for 5 min, the supernatant was transferred to a new 1.5 mL Eppendorf tube, and the digest was dried in a Speedvac and stored at -20°C until MS analysis.

Proximity labeling with TurboID

Propyzamide was added to transgenic *Arabidopsis* cell cultures, and 2 hours later 50 μM biotin was added for 2 hours at 25°C. To block proteasomal degradation, 20 μM MG132 was added simultaneously with biotin to one half of the cells. Cells treated for 4 hours with propyzamide, were harvested as previously described (Van Leene et al., 2015). 9 g harvested cells were ground to homogeneity in liquid nitrogen and added to 6 mL of extraction buffer (100mM Tris pH7.5, 2% SDS, 8 M urea). Cells were mechanically disrupted by three repetitive freeze-thaw cycles, freezing in liquid nitrogen for 5 min and thawing in tap water for 25 min. Samples were subsequently sonicated (25 sec ON, 35 sec OFF, 25 sec ON) and incubated for 1 hour under gentle rotation at room temperature. Supernatants were separated from cell debris by two consecutive centrifugation steps at 20000 rpm for 20 min at room temperature. The extracts were passed through a GF-prefilter and 0.45-μm filter (Sartorius AG) and excess of free biotin was removed on a PD 10 Desalting Columns (Merck), which was first equilibrated with binding buffer (100 mM Tris pH7.5, 2% SDS, 7.5 M urea). Proteins were eluted from the PD 10 Desalting Columns with 3.5 mL extraction buffer. Extracts were divided in 3 experimental repeats and incubated overnight at room temperature with 100 μL Streptavidin Sepharose High Performance beads (Amersham). Supernatant was removed by centrifugation at 1500 rpm for 1 min and subsequent transfer to a mobicol column (Mo Bi Tec). Beads were washed with 4 mL of binding buffer, 800 μL of high salt buffer (1 M NaCl, 100 mM Tris-HCl pH7.5) for 30 min and 2x 800 μL of ultrapure water. On-bead digestion was preceded with a washing step of 3.2 mL 50 mM ammonium bicarbonate pH8.0. Beads were resolved in 200 μl of 50 mM ammonium bicarbonate pH8.0 and 4 μl Trypsin/LysC mix (Promega)

was added followed by incubated at 37°C overnight with agitation. An additional 2 µL Trypsin/LysC mix was added for another 2 hours at 37°C. After centrifugation at 1500 rpm for 1 min, the digest was transferred to an Eppendorf and beads were washed with 150 µL of HPLC grade water. Digest and wash sample were pooled together. Biotinylated peptides were eluted from the beads in two consecutive elution step in 300 µL solution of 0.2 % TFA, 0.1% FA and 80 % acetonitrile. Eluted peptides were lyophilized and resolved in the trypsin digest. Peptide samples were desalted with C18 Omix tips (Agilent), which were first equilibrated with a pre-wash buffer (80% acetonitrile (AcN), 0.1% FA) and wash buffer (0.1% FA). Peptides were loaded on the C18 matrix and washed with wash buffer. Peptides were eluted in elution buffer (60% acetonitrile (AcN), 0.1% FA) and lyophilized for LC-MS/MS analysis.

Mass Spectrometry and data analysis

On-bead digested samples were analyzed on a Q Exactive (ThermoFisher Scientific) and co-purified proteins were identified using standard procedures (Van Leene et al., 2015). After identification of AP-MS co-purified proteins, the protein list was filtered versus a list of non-specific proteins, assembled similarly as described (Van Leene et al., 2015). In this case, the list of non-specific proteins was built from 123 AP-MS experiments, covering 13 baitgroups. All proteins appearing with 3 or more baitgroups were included in the list of non-specific proteins, resulting in a background list of 3186 proteins. For each bait to be analyzed, the background was marked in the list of the co-purified proteins. To prevent true interactors being filtered out because of their presence in the list of non-specific proteins, a T-test was performed using the average normalized spectral abundance factors (NSAF) of the identified proteins in the bait pull-downs versus the corresponding average NSAF in a control set of pull-downs with non-related baits (other CCS52 experiments were removed). Proteins identified with at least two peptides in at least two experiments, that were not present in the background list or showed high (at least 10-fold) and significant [-log10(p-value(T-test)) ≥10] enrichment versus a large dataset of pull-downs with non-related bait proteins, were retained. The proteins present in the background list without high and significant enrichment versus the large dataset were removed.

For quantitative identification of specific interactors compared to propyzamide synchronized cultures, MaxQuant LFQ values were analyzed in Perseus as follows. First, LFQ values were Log2 transformed and potential contaminants and reverse hits were removed. Replicates were then grouped and proteins were filtered for at least two identifications in at least one group. Missing values were imputed from a normal distribution around the detection limit per sample (width = 0.3 and down shift = 1.8). Triplicate CCS52A1, CCS52A2 or CCS52B pull-downs on time point zero (T0) were analyzed against six AUR1 and AUR2 T0 pull-downs. Triplicate CCS52A1, CCS52A2 or CCS52B pull-downs on time point ten (T10) were analyzed against triplicate wild type PSB-D T10 pull-downs. Significantly enriched proteins were identified through a volcano plot analysis with thresholds FDR = 0.0001 (T0) or FDR = 0.001 (T10) and S0 = 1.

Trypsin peptides from TurboID-CCS52B proximity labeling experiments were analyzed on a Q Exactive (ThermoFisher Scientific) as previously reported (Nelissen et al., 2015). After MS-based identification of co-purified proteins, the triplicate TurboID experiments were grouped and specific proteins were detected by comparison of MaxQuant LFQ values against a data set consisting of 66 TurboID experiments on unrelated baits, leaving out endocytosis related experiments (baits: TPLATE, AP2M and TML). First, LFQ values were Log2 transformed and potential contaminants and reverse hits were removed. Replicates were then grouped and proteins were filtered for at least two identifications in at least one group. Missing values were imputed from a normal distribution around the detection limit per sample (width = 0.3 and down shift = 1.8). Significantly enriched proteins were identified through a volcano plot analysis with thresholds FDR = 0.05, S0 = 0.5.

For quantitative identification of APC12 (Q8H1U3) in the AP-MS data, identified proteins for both time points per bait were grouped and MaxQuant LFQ values were analyzed in Perseus against a control data set containing 9 pull-downs on 3 baits (PSB-D, AUR1 and AUR2) in propyzamide synchronized cultures. For quantitative identification of APC12 (Q8H1U3) in the TurboID experiments, MaxQuant LFQ values were analyzed in Perseus against a data set consisting of 96 TurboID experiments. As database Araport11 was used with addition of the UniProt proteins sequence of APC12 (Q8H1U3).

References

Aguilar-Hernandez V, Kim DY, Stankey RJ, Scalf M, Smith LM, Vierstra RD (2017) Mass Spectrometric Analyses Reveal a Central Role for Ubiquitylation in Remodeling the Arabidopsis Proteome during Photomorphogenesis. Molecular Plant **10:** 846-865

Akhter S, Richie CT, Deng JM, Brey E, Zhang X, Patrick C, Jr., Behringer RR, Legerski RJ (2004) Deficiency in SNM1 abolishes an early mitotic checkpoint induced by spindle stress. Mol Cell Biol **24:** 10448-10455

Alfieri C, Zhang S, Barford D (2017) Visualizing the complex functions and mechanisms of the anaphase promoting complex/cyclosome (APC/C). Open Biol **7**

Amador V, Ge S, Santamaria PG, Guardavaccaro D, Pagano M (2007) APC/C(Cdc20) controls the ubiquitin-mediated degradation of p21 in prometaphase. Mol Cell **27:** 462-473

Arlot-Bonnemains Y, Klotzbucher A, Giet R, Uzbekov R, Bihan R, Prigent C (2001) Identification of a functional destruction box in the Xenopus laevis aurora-A kinase pEg2. FEBS Lett **508:** 149-152

Bar M, Aharon M, Benjamin S, Rotblat B, Horowitz M, Avni A (2008) AtEHDs, novel Arabidopsis EH-domain-containing proteins involved in endocytosis. Plant J **55:** 1025-1038

Beck M, Komis G, Ziemann A, Menzel D, Samaj J (2011) Mitogen-activated protein kinase 4 is involved in the regulation of mitotic and cytokinetic microtubule transitions in Arabidopsis thaliana. New Phytol **189:** 1069-1083

Berson T, von Wangenheim D, Takac T, Samajova O, Rosero A, Ovecka M, Komis G, Stelzer EH, Samaj J (2014) Trans-Golgi network localized small GTPase RabA1d is involved in cell plate formation and oscillatory root hair growth. BMC Plant Biol **14:** 252

Blilou I, Frugier F, Folmer S, Serralbo O, Willemsen V, Wolkenfelt H, Eloy NB, Ferreira PC, Weisbeek P, Scheres B (2002) The Arabidopsis HOBBIT gene encodes a CDC27 homolog that links the plant cell cycle to progression of cell differentiation. Genes Dev **16:** 2566-2575

Boruc J, Inze D, Russinova E (2010) A high-throughput bimolecular fluorescence complementation protein-protein interaction screen identifies functional Arabidopsis CDKA/B-CYCD4/5 complexes. Plant Signal Behav **5:** 1276-1281

Boruc J, Mylle E, Duda M, De Clercq R, Rombauts S, Geelen D, Hilson P, Inze D, Van Damme D, Russinova E (2010) Systematic localization of the Arabidopsis core cell cycle proteins reveals novel cell division complexes. Plant Physiol **152:** 553-565

Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inze D, De Veylder L, Russinova E (2010) Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network. Plant Cell **22:** 1264-1280

Boudolf V, Lammens T, Boruc J, Van Leene J, Van Den Daele H, Maes S, Van Isterdael G, Russinova E, Kondorosi E, Witters E, De Jaeger G, Inze D, De Veylder L (2009) CDKB1;1 forms a functional complex with CYCA2;3 to suppress endocycle onset. Plant Physiol **150:** 1482-1493

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Brembu T, Winge P, Seem M, Bones AM (2004) NAPP and PIRP encode subunits of a putative wave regulatory protein complex involved in plant cell morphogenesis. Plant Cell **16:** 2335-2349

Buschmann H, Muller S (2019) Update on plant cytokinesis: rule and divide. Curr Opin Plant Biol **52:** 97-105

Capron A, Serralbo O, Fulop K, Frugier F, Parmentier Y, Dong A, Lecureuil A, Guerche P, Kondorosi E, Scheres B, Genschik P (2003) The Arabidopsis anaphase-promoting complex or cyclosome: molecular and genetic characterization of the APC2 subunit. Plant Cell **15:** 2370-2382

Chen P, Takatsuka H, Takahashi N, Kurata R, Fukao Y, Kobayashi K, Ito M, Umeda M (2017) Arabidopsis R1R2R3-Myb proteins are essential for inhibiting cell division in response to DNA damage. Nat Commun **8:** 635

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J **89:** 789-804

Coyaud E, Mis M, Laurent EM, Dunham WH, Couzens AL, Robitaille M, Gingras AC, Angers S, Raught B (2015) BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates. Mol Cell Proteomics **14:** 1781-1795

Criqui MC, de Almeida Engler J, Camasses A, Capron A, Parmentier Y, Inze D, Genschik P (2002) Molecular characterization of plant ubiquitin-conjugating enzymes belonging to the UbcP4/E2-C/UBCx/UbcH10 gene family. Plant Physiol **130:** 1230-1240

Cromer L, Heyman J, Touati S, Harashima H, Araou E, Girard C, Horlow C, Wassmann K, Schnittger A, De Veylder L, Mercier R (2012) OSD1 promotes meiotic progression via APC/C inhibition and forms a regulatory network with TDM and CYCA1;2/TAM. PLoS Genet **8:** e1002865

Cromer L, Jolivet S, Singh DK, Berthier F, De Winne N, De Jaeger G, Komaki S, Prusicki MA, Schnittger A, Guerois R, Mercier R (2019) Patronus is the elusive plant securin, preventing chromosome separation by antagonizing separase. Proc Natl Acad Sci U S A **116:** 16018-16027

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Simon M, Jenczewski E, Mercier R (2008) Mutations in AtPS1 (Arabidopsis thaliana parallel spindle 1) lead to the production of diploid pollen grains. PLoS Genet **4:** e1000274

da Fonseca PC, Kong EH, Zhang Z, Schreiber A, Williams MA, Morris EP, Barford D (2011) Structures of APC/C(Cdh1) with substrates identify Cdh1 and Apc10 as the D-box co-receptor. Nature **470:** 274-278

Davey NE, Morgan DO (2016) Building a Regulatory Network with Short Linear Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting Complex. Mol Cell **64:** 12-23

Demidov D, Van Damme D, Geelen D, Blattner FR, Houben A (2005) Identification and dynamics of two classes of aurora-like kinases in Arabidopsis and other plants. Plant Cell **17:** 836-848

den Elzen N, Pines J (2001) Cyclin a is destroyed in prometaphase and can delay chromosome alignment and anaphase. Journal of Cell Biology **153:** 121-135

Di Fiore B, Davey NE, Hagting A, Izawa D, Mansfeld J, Gibson TJ, Pines J (2015) The ABBA motif binds APC/C activators and is shared by APC/C substrates and regulators. Dev Cell **32:** 358-372

Doerr A (2018) Proximity labeling with TurboID. Nat Methods **15:** 764

Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GT, Inze D, Ferreira PC (2011) The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. Plant J **68:** 351-363

Eloy NB, Gonzalez N, Van Leene J, Maleux K, Vanhaeren H, De Milde L, Dhondt S, Vercruysse L, Witters E, Mercier R, Cromer L, Beemster GTS, Remaut H, Van Montagu MCE, De Jaeger G, Ferreira PCG, Inze D (2012) SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization. Proceedings of the National Academy of Sciences of the United States of America **109:** 13853-13858

Eloy NB, Lima MD, Ferreira PCG, Inze D (2015) The Role of the Anaphase-Promoting Complex/Cyclosome in Plant Growth. Critical Reviews in Plant Sciences **34:** 487-505

Floyd S, Pines J, Lindon C (2008) APC/C Cdh1 targets aurora kinase to control reorganization of the mitotic spindle at anaphase. Curr Biol **18:** 1649-1658

Fulop K, Tarayre S, Kelemen Z, Horvath G, Kevei Z, Nikovics K, Bako L, Brown S, Kondorosi A, Kondorosi E (2005) Arabidopsis anaphase-promoting complexes: multiple activators and wide range of substrates might keep APC perpetually busy. Cell Cycle **4:** 1084-1092

Ganguly A, Bhattacharya R, Cabral F (2008) Cell cycle dependent degradation of MCAK: evidence against a role in anaphase chromosome movement. Cell Cycle **7:** 3187-3193

Garnett MJ, Mansfeld J, Godwin C, Matsusaka T, Wu J, Russell P, Pines J, Venkitaraman AR (2009) UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. Nat Cell Biol **11:** 1363-1369

Garrido G, Vernos I (2016) Non-centrosomal TPX2-Dependent Regulation of the Aurora A Kinase: Functional Implications for Healthy and Pathological Cell Division. Front Oncol **6:** 88

Gingras AC, Abe KT, Raught B (2019) Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. Curr Opin Chem Biol **48:** 44-54

Glotzer M, Murray AW, Kirschner MW (1991) Cyclin is degraded by the ubiquitin pathway. Nature **349:** 132-138

Gmachl M, Gieffers C, Podtelejnikov AV, Mann M, Peters JM (2000) The RING-H2 finger protein APC11 and the E2 enzyme UBC4 are sufficient to ubiquitinate substrates of the anaphase-promoting complex. Proc Natl Acad Sci U S A **97:** 8973-8978

Gorr IH, Boos D, Stemmann O (2005) Mutual inhibition of separase and Cdk1 by two-step complex formation. Mol Cell **19:** 135-141

Haga N, Kato K, Murase M, Araki S, Kubo M, Demura T, Suzuki K, Muller I, Voss U, Jurgens G, Ito M (2007) R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in Arabidopsis thaliana. Development **134:** 1101-1110

Hamperl S, Brown CR, Perez-Fernandez J, Huber K, Wittner M, Babl V, Stockl U, Boeger H, Tschochner H, Milkereit P, Griesenbeck J (2014) Purification of specific chromatin domains from single-copy gene loci in Saccharomyces cerevisiae. Methods Mol Biol **1094:** 329-341

Hellmuth S, Bottger F, Pan C, Mann M, Stemmann O (2014) PP2A delays APC/C-dependent degradation of separase-associated but not free securin. EMBO J **33:** 1134-1147

Heyman J, Cools T, Vandenbussche F, Heyndrickx KS, Van Leene J, Vercauteren I, Vanderauwera S, Vandepoele K, De Jaeger G, Van Der Straeten D, De Veylder L (2013) ERF115 controls root quiescent center cell division and stem cell replenishment. Science **342:** 860-863

Heyman J, De Veylder L (2012) The anaphase-promoting complex/cyclosome in control of plant development. Mol Plant **5:** 1182-1194

Heyman J, Van den Daele H, De Wit K, Boudolf V, Berckmans B, Verkest A, Alvim Kamei CL, De Jaeger G, Koncz C, De Veylder L (2011) Arabidopsis ULTRAVIOLET-B-INSENSITIVE4 maintains cell division activity by temporal inhibition of the anaphase-promoting complex/cyclosome. Plant Cell **23:** 4394-4410

Horn SR, Thomenius MJ, Johnson ES, Freel CD, Wu JQ, Coloff JL, Yang CS, Tang W, An J, Ilkayeva OR, Rathmell JC, Newgard CB, Kornbluth S (2011) Regulation of mitochondrial morphology by APC/CCdh1-mediated control of Drp1 stability. Mol Biol Cell **22:** 1207-1216

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

Iconomou M, Saunders DN (2016) Systematic approaches to identify E3 ligase substrates. Biochem J **473:** 4083-4101

Iwata E, Ikeda S, Matsunaga S, Kurata M, Yoshioka Y, Criqui MC, Genschik P, Ito M (2011) GIGAS CELL1, a novel negative regulator of the anaphase-promoting complex/cyclosome, is required for proper mitotic progression and cell fate determination in Arabidopsis. Plant Cell **23:** 4382-4393

Kaspar M, Dienemann A, Schulze C, Sprenger F (2001) Mitotic degradation of cyclin A is mediated by multiple and novel destruction signals. Curr Biol **11:** 685-690

Kataria M, Yamano H (2019) Interplay between Phosphatases and the Anaphase-Promoting Complex/Cyclosome in Mitosis. Cells **8**

Kawabe A, Matsunaga S, Nakagawa K, Kurihara D, Yoneda A, Hasezawa S, Uchiyama S, Fukui K (2005) Characterization of plant Aurora kinases during mitosis. Plant Mol Biol **58:** 1-13

Kevei Z, Baloban M, Da Ines O, Tiricz H, Kroll A, Regulski K, Mergaert P, Kondorosi E (2011) Conserved CDC20 cell cycle functions are carried out by two of the five isoforms in Arabidopsis thaliana. PLoS One **6:** e20618

Kim DY, Scalf M, Smith LM, Vierstra RD (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. Plant Cell **25:** 1523-1540

Kosetsu K, Matsunaga S, Nakagami H, Colcombet J, Sasabe M, Soyano T, Takahashi Y, Hirt H, Machida Y (2010) The MAP kinase MPK4 is required for cytokinesis in Arabidopsis thaliana. Plant Cell **22:** 3778-3790

Kraft C, Herzog F, Gieffers C, Mechtler K, Hagting A, Pines J, Peters JM (2003) Mitotic regulation of the human anaphase-promoting complex by phosphorylation. EMBO J **22:** 6598-6609

Kramer ER, Scheuringer N, Podtelejnikov AV, Mann M, Peters JM (2000) Mitotic regulation of the APC activator proteins CDC20 and CDH1. Mol Biol Cell **11:** 1555-1569

Kurasawa Y, Earnshaw WC, Mochizuki Y, Dohmae N, Todokoro K (2004) Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. EMBO J **23:** 3237-3248

Kwee HS, Sundaresan V (2003) The NOMEGA gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in Arabidopsis. Plant J **36:** 853-866

Lammens T, Boudolf V, Kheibarshekan L, Zalmas LP, Gaamouche T, Maes S, Vanstraelen M, Kondorosi E, La Thangue NB, Govaerts W, Inze D, De Veylder L (2008) Atypical E2F activity restrains APC/CCCS52A2 function obligatory for endocycle onset. Proc Natl Acad Sci U S A **105:** 14721-14726

Lee SJ, Rodriguez-Bravo V, Kim H, Datta S, Foley EA (2017) The PP2A(B56) phosphatase promotes the association of Cdc20 with APC/C in mitosis. J Cell Sci **130:** 1760-1771

Lin Q, Wang D, Dong H, Gu S, Cheng Z, Gong J, Qin R, Jiang L, Li G, Wang JL, Wu F, Guo X, Zhang X, Lei C, Wang H, Wan J (2012) Rice APC/C(TE) controls tillering by mediating the degradation of MONOCULM 1. Nat Commun **3:** 752

Lin Q, Wu F, Sheng P, Zhang Z, Zhang X, Guo X, Wang J, Cheng Z, Wang J, Wang H, Wan J (2015) The SnRK2-APC/C(TE) regulatory module mediates the antagonistic action of gibberellic acid and abscisic acid pathways. Nat Commun **6:** 7981

Linares JF, Amanchy R, Greis K, Diaz-Meco MT, Moscat J (2011) Phosphorylation of p62 by cdk1 controls the timely transit of cells through mitosis and tumor cell proliferation. Mol Cell Biol **31:** 105-117

Littlepage LE, Ruderman JV (2002) Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev **16:** 2274-2285

Liu ST, Zhang H (2016) The mitotic checkpoint complex (MCC): looking back and forth after 15 years. AIMS Mol Sci **3:** 597-634

Lorenzo-Orts L, Witthoeft J, Deforges J, Martinez J, Loubery S, Placzek A, Poirier Y, Hothorn LA, Jaillais Y, Hothorn M (2019) Concerted expression of a cell cycle regulator and a metabolic enzyme from a bicistronic transcript in plants. Nat Plants **5:** 184-193

Mansfeld J, Collin P, Collins MO, Choudhary JS, Pines J (2011) APC15 drives the turnover of MCC-CDC20 to make the spindle assembly checkpoint responsive to kinetochore attachment. Nat Cell Biol **13:** 1234-1243

Mao G, Chan J, Calder G, Doonan JH, Lloyd CW (2005) Modulated targeting of GFP-AtMAP65-1 to central spindle microtubules during division. Plant J **43:** 469-478

Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics **6:** 601-610

Marc J, Granger CL, Brincat J, Fisher DD, Kao T, McCubbin AG, Cyr RJ (1998) A GFP-MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells. Plant Cell **10:** 1927-1940

Martinsson-Ahlzen HS, Liberal V, Grunenfelder B, Chaves SR, Spruck CH, Reed SI (2008) Cyclin-dependent kinase-associated proteins Cks1 and Cks2 are essential during early embryogenesis and for cell cycle progression in somatic cells. Mol Cell Biol **28:** 5698-5709

Mathe C, Garda T, Freytag C, M MH (2019) The Role of Serine-Threonine Protein Phosphatase PP2A in Plant Oxidative Stress Signaling-Facts and Hypotheses. Int J Mol Sci **20**

Mathieu-Rivet E, Gevaudant F, Sicard A, Salar S, Do PT, Mouras A, Fernie AR, Gibon Y, Rothan C, Chevalier C, Hernould M (2010) Functional analysis of the anaphase promoting complex activator CCS52A highlights the crucial role of endo-reduplication for fruit growth in tomato. Plant J **62:** 727-741

Matyskiela ME, Morgan DO (2009) Analysis of activator-binding sites on the APC/C supports a cooperative substrate-binding mechanism. Mol Cell **34:** 68-80

Mazumdar M, Sundareshan S, Misteli T (2004) Human chromokinesin KIF4A functions in chromosome condensation and segregation. J Cell Biol **166:** 613-620

Menges M, de Jager SM, Gruissem W, Murray JA (2005) Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. Plant J **41:** 546-566

Menges M, Hennig L, Gruissem W, Murray JA (2003) Genome-wide gene expression in an Arabidopsis cell suspension. Plant Mol Biol **53:** 423-442

Menges M, Murray JA (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J **30:** 203-212

Miyagishima SY, Kuwayama H, Urushihara H, Nakanishi H (2008) Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. Proc Natl Acad Sci U S A **105:** 15202-15207

Nagata T, Kumagai F (1999) Plant cell biology through the window of the highly synchronized tobacco BY-2 cell line. Methods Cell Sci **21:** 123-127

Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco by-2 Cell-Line as the Hela-Cell in the Cell Biology of Higher-Plants. International Review of Cytology-a Survey of Cell Biology **132:** 1-30

Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, Van Lijsebettens M, Goormachtig S, Vandepoele K, Van Leene J, Muszynski M, Gevaert K, Inze D, De Jaeger G (2015) Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf. Plant Cell **27:** 1605-1619

Patra D, Dunphy WG (1998) Xe-p9, a Xenopus Suc1/Cks protein, is essential for the Cdc2-dependent phosphorylation of the anaphase- promoting complex at mitosis. Genes Dev **12:** 2549-2559

Pfleger CM, Kirschner MW (2000) The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev **14:** 655-665

Planchais S, Glab N, Inze D, Bergounioux C (2000) Chemical inhibitors: a tool for plant cell cycle studies. FEBS Lett **476:** 78-83

Reinecke JB, Katafiasz D, Naslavsky N, Caplan S (2015) Novel functions for the endocytic regulatory proteins MICAL-L1 and EHD1 in mitosis. Traffic **16:** 48-67

Rojas CA, Eloy NB, Lima Mde F, Rodrigues RL, Franco LO, Himanen K, Beemster GT, Hemerly AS, Ferreira PC (2009) Overexpression of the Arabidopsis anaphase promoting complex subunit CDC27a increases growth rate and organ size. Plant Mol Biol **71:** 307-318

Rossio V, Michimoto T, Sasaki T, Ohbayashi I, Kikuchi Y, Yoshida S (2013) Nuclear PP2A-Cdc55 prevents APC-Cdc20 activation during the spindle assembly checkpoint. J Cell Sci **126:** 4396-4405

Saracco SA, Hansson M, Scalf M, Walker JM, Smith LM, Vierstra RD (2009) Tandem affinity purification and mass spectrometric analysis of ubiquitylated proteins in Arabidopsis. Plant J **59:** 344-358

Schwickart M, Havlis J, Habermann B, Bogdanova A, Camasses A, Oelschlaegel T, Shevchenko A, Zachariae W (2004) Swm1/Apc13 is an evolutionarily conserved subunit of the anaphase-promoting complex stabilizing the association of Cdc16 and Cdc27. Mol Cell Biol **24:** 3562-3576

Seshacharyulu P, Pandey P, Datta K, Batra SK (2013) Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. Cancer Lett **335:** 9-18

Singh SA, Winter D, Kirchner M, Chauhan R, Ahmed S, Ozlu N, Tzur A, Steen JA, Steen H (2014) Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J **33:** 385-399

Smalle J, Vierstra RD (2004) The ubiquitin 26S proteasome proteolytic pathway. Annu Rev Plant Biol **55:** 555-590

Smet W, Sevilem I, de Luis Balaguer MA, Wybouw B, Mor E, Miyashima S, Blob B, Roszak P, Jacobs TB, Boekschoten M, Hooiveld G, Sozzani R, Helariutta Y, De Rybel B (2019) DOF2.1 Controls Cytokinin-Dependent Vascular Cell Proliferation Downstream of TMO5/LHW. Curr Biol **29:** 520-529 e526

Song MS, Carracedo A, Salmena L, Song SJ, Egia A, Malumbres M, Pandolfi PP (2011) Nuclear PTEN regulates the APC-CDH1 tumor-suppressive complex in a phosphatase-independent manner. Cell **144:** 187-199

Spinner L, Gadeyne A, Belcram K, Goussot M, Moison M, Duroc Y, Eeckhout D, De Winne N, Schaefer E, Van De Slijke E, Persiau G, Witters E, Gevaert K, De Jaeger G, Bouchez D, Van Damme D, Pastuglia M (2013) A protein phosphatase 2A complex spatially controls plant cell division. Nat Commun **4:** 1863

Spruck CH, de Miguel MP, Smith AP, Ryan A, Stein P, Schultz RM, Lincoln AJ, Donovan PJ, Reed SI (2003) Requirement of Cks2 for the first metaphase/anaphase transition of mammalian meiosis. Science **300:** 647-650

Stewart S, Fang G (2005) Anaphase-promoting complex/cyclosome controls the stability of TPX2 during mitotic exit. Mol Cell Biol **25:** 10516-10527

Stewart S, Fang G (2005) Destruction box-dependent degradation of aurora B is mediated by the anaphase-promoting complex/cyclosome and Cdh1. Cancer Res **65:** 8730-8735

Sullivan JA, Shirasu K, Deng XW (2003) The diverse roles of ubiquitin and the 26S proteasome in the life of plants. Nat Rev Genet **4:** 948-958

Suzuki T, Nakajima S, Inagaki S, Hirano-Nakakita M, Matsuoka K, Demura T, Fukuda H, Morikami A, Nakamura K (2005) TONSOKU is expressed in S phase of the cell cycle and its defect delays cell cycle progression in Arabidopsis. Plant Cell Physiol **46:** 736-742

Svenning S, Lamark T, Krause K, Johansen T (2011) Plant NBR1 is a selective autophagy substrate and a functional hybrid of the mammalian autophagic adapters NBR1 and p62/SQSTM1. Autophagy **7:** 993-1010

Tang Z, Li B, Bharadwaj R, Zhu H, Ozkan E, Hakala K, Deisenhofer J, Yu H (2001) APC2 Cullin protein and APC11 RING protein comprise the minimal ubiquitin ligase module of the anaphase-promoting complex. Mol Biol Cell **12:** 3839-3851

Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc **11:** 2301-2319

Uzunova K, Dye BT, Schutz H, Ladurner R, Petzold G, Toyoda Y, Jarvis MA, Brown NG, Poser I, Novatchkova M, Mechtler K, Hyman AA, Stark H, Schulman BA, Peters JM (2012) APC15 mediates CDC20 autoubiquitylation by APC/C(MCC) and disassembly of the mitotic checkpoint complex. Nat Struct Mol Biol **19:** 1116-1123

Van Damme D, Bouget FY, Van Poucke K, Inze D, Geelen D (2004) Molecular dissection of plant cytokinesis and phragmoplast structure: a survey of GFP-tagged proteins. Plant J **40:** 386-398

Van Damme D, Coutuer S, De Rycke R, Bouget FY, Inze D, Geelen D (2006) Somatic cytokinesis and pollen maturation in Arabidopsis depend on TPLATE, which has domains similar to coat proteins. Plant Cell **18:** 3502-3518

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, Martens L, Witters E, Gevaert K, De Jaeger G (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc **10:** 169-187

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, De Bodt S, Maere S, Laukens K, Pharazyn A, Ferreira PC, Eloy N, Renne

C, Meyer C, Faure JD, Steinbrenner J, Beynon J, Larkin JC, Van de Peer Y, Hilson P, Kuiper M, De Veylder L, Van Onckelen H, Inze D, Witters E, De Jaeger G (2010) Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana. Mol Syst Biol **6:** 397

Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, Henderickx K, De Vijlder T, Abdelkrim A, Pharazyn A, Van Onckelen H, Inze D, Witters E, De Jaeger G (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics **6:** 1226-1238

van Zon W, Ogink J, ter Riet B, Medema RH, te Riele H, Wolthuis RM (2010) The APC/C recruits cyclin B1-Cdk1-Cks in prometaphase before D box recognition to control mitotic exit. J Cell Biol **190:** 587-602

Vanstraelen M, Baloban M, Da Ines O, Cultrone A, Lammens T, Boudolf V, Brown SC, De Veylder L, Mergaert P, Kondorosi E (2009) APC/C-CCS52A complexes control meristem maintenance in the Arabidopsis root. Proc Natl Acad Sci U S A **106:** 11806-11811

Vanstraelen M, Inze D, Geelen D (2006) Mitosis-specific kinesins in Arabidopsis. Trends Plant Sci **11:** 167-175

Verhoeven HA, Sree Ramulu K, Dijkhuis P (1990) A comparison of the effects of various spindle toxins on metaphase arrest and formation of micronuclei in cell-suspension cultures ofNicotiana plumbaginifolia. Planta **182:** 408-414

Vos JW, Pieuchot L, Evrard JL, Janski N, Bergdoll M, de Ronde D, Perez LH, Sardon T, Vernos I, Schmit AC (2008) The plant TPX2 protein regulates prospindle assembly before nuclear envelope breakdown. Plant Cell **20:** 2783-2797

Walton A, Stes E, Cybulski N, Van Bel M, Inigo S, Durand AN, Timmerman E, Heyman J, Pauwels L, De Veylder L, Goossens A, De Smet I, Coppens F, Goormachtig S, Gevaert K (2016) It's Time for Some "Site"-Seeing: Novel Tools to Monitor the Ubiquitin Landscape in Arabidopsis thaliana. Plant Cell **28:** 6-16

Wang F, Zhu D, Huang X, Li S, Gong Y, Yao Q, Fu X, Fan LM, Deng XW (2009) Biochemical insights on degradation of Arabidopsis DELLA proteins gained from a cell-free assay system. Plant Cell **21:** 2378-2390

Wang Y, Hou Y, Gu H, Kang D, Chen Z, Liu J, Qu LJ (2012) The Arabidopsis APC4 subunit of the anaphase-promoting complex/cyclosome (APC/C) is critical for both female gametogenesis and embryogenesis. Plant J **69:** 227-240

Wang Y, Hou Y, Gu H, Kang D, Chen ZL, Liu J, Qu LJ (2013) The Arabidopsis anaphase-promoting complex/cyclosome subunit 1 is critical for both female gametogenesis and embryogenesis(F). J Integr Plant Biol **55:** 64-74

Willems P, Horne A, Van Parys T, Goormachtig S, De Smet I, Botzki A, Van Breusegem F, Gevaert K (2019) The Plant PTM Viewer, a central resource for exploring plant protein modifications. Plant Journal **99:** 752-762

Wlodarchak N, Xing Y (2016) PP2A as a master regulator of the cell cycle. Crit Rev Biochem Mol Biol **51:** 162-184

Wolthuis R, Clay-Farrace L, van Zon W, Yekezare M, Koop L, Ogink J, Medema R, Pines J (2008) Cdc20 and Cks direct the spindle checkpoint-independent destruction of cyclin A. Mol Cell **30:** 290-302

Xu G, Jaffrey SR (2013) Proteomic identification of protein ubiquitination events. Biotechnol Genet Eng Rev **29:** 73-109

Yamano H (2019) APC/C: current understanding and future perspectives. F1000Res **8**

Yang W, Wightman R, Meyerowitz EM (2017) Cell Cycle Control by Nuclear Sequestration of CDC20 and CDH1 mRNA in Plant Stem Cells. Mol Cell **68:** 1108-1119 e1103

Zhang S, Chang L, Alfieri C, Zhang Z, Yang J, Maslen S, Skehel M, Barford D (2016) Molecular mechanism of APC/C activation by mitotic phosphorylation. Nature **533:** 260-264

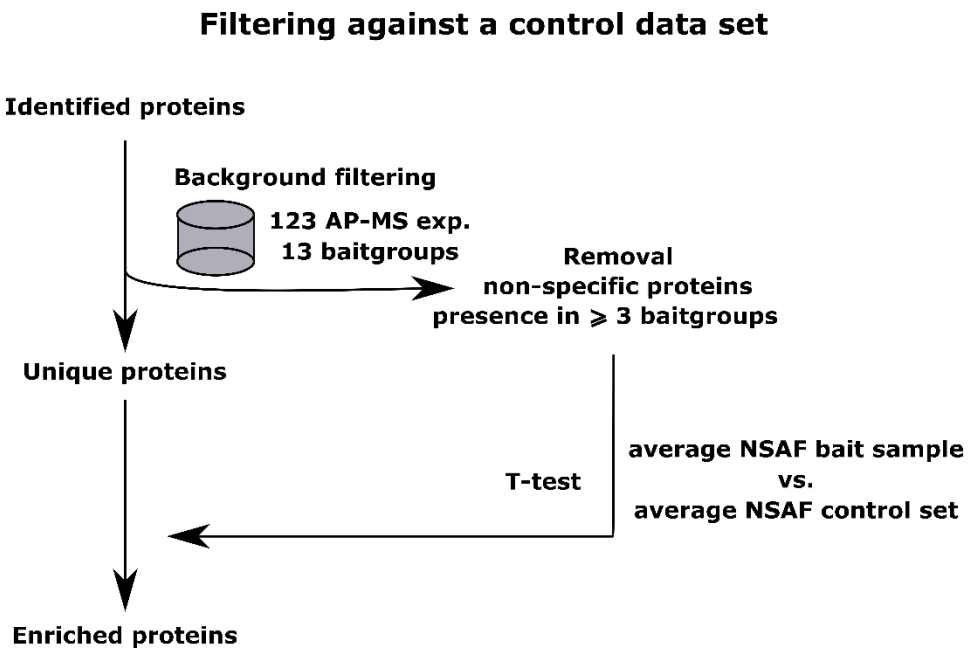
Zhou Y, Tergemina E, Cui H, Forderer A, Hartwig B, Velikkakam James G, Schneeberger K, Turck F (2017) Ctf4-related protein recruits LHP1-PRC2 to maintain H3K27me3 levels in dividing cells in Arabidopsis thaliana. Proc Natl Acad Sci U S A **114:** 4833-4838

Zhou Z, He M, Shah AA, Wan Y (2016) Insights into APC/C: from cellular function to diseases and therapeutics. Cell Div **11:** 9

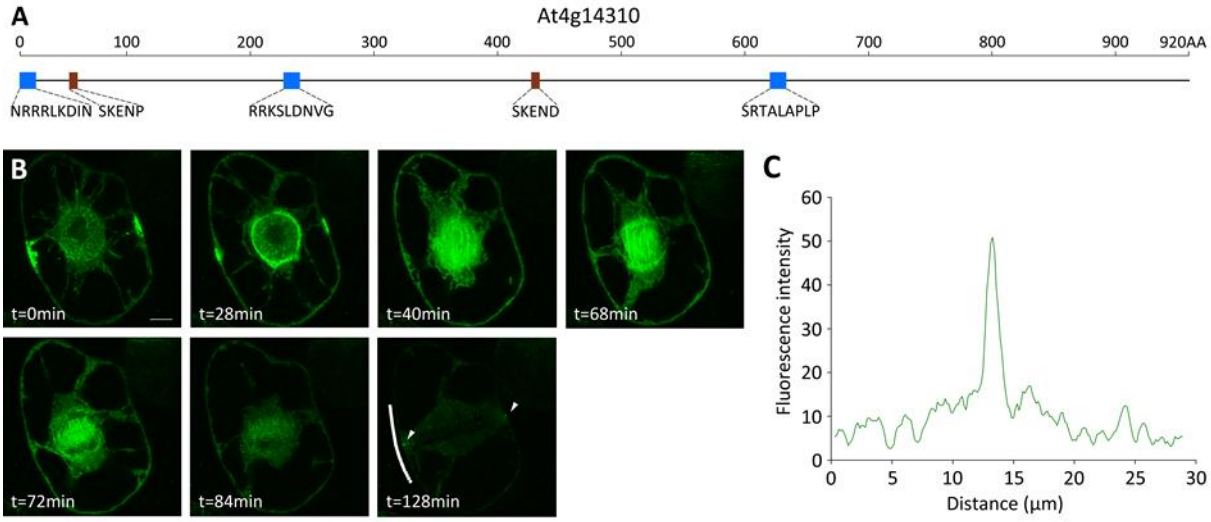
Zhu C, Jiang W (2005) Cell cycle-dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. Proc Natl Acad Sci U S A **102:** 343-348

Zhu C, Zhao J, Bibikova M, Leverson JD, Bossy-Wetzel E, Fan JB, Abraham RT, Jiang W (2005) Functional analysis of human microtubule-based motor proteins, the kinesins and dyneins, in mitosis/cytokinesis using RNA interference. Mol Biol Cell **16:** 3187-3199

Supplementary data



Supplementary figure 1: Schematic overview of the different filtering steps applied to identify enriched proteins after AP-MS on the CCS52 co-activators. After identification of AP-MS co-purified proteins, the protein list was filtered versus a list of non-specific proteins built from 123 AP-MS experiments, covering 13 baitgroups. All proteins appearing with 3 or more baitgroups were included in the list of non-specific proteins. To prevent true interactors being filtered out because of their presence in the list of non-specific proteins, a T-test was performed using the average normalized spectral abundance factors (NSAF) of the identified proteins in the bait pull-downs versus the corresponding average NSAF in a control set of pull-downs with non-related baits (other CCS52 experiments were removed). Proteins identified with at least two peptides in at least two experiments, that were not present in the background list or showed high (at least 10-fold) and significant [-log10(p-value(T-test)) ≥10] enrichment versus a large dataset of pull-downs with non-related bait proteins, were retained.



Supplementary figure 2: AT4G14310 is a novel microtubule-associated protein. **A)** Protein structure of AT4G14310 with indication of destruction boxes (D-box, blue; KEN-box, red) and their respective sequence motif. **B)** Time lapse of a dividing tobacco BY-2 cell expressing 35S::GFP-AT4G14310. AT4G14310 associates with the preprophase band, the spindle and early phragmoplast. Upon cell plate insertion, the GFP-AT4G14310 signal reappears at the exact insertion site (arrowheads). **C)** GFP fluorescence intensity profile of the last frame from the time lapse in panel B along the plasma membrane (white line), indicating GFP-AT4G14310 fluorescence at the cell plate insertion site. (Scale bar, 10µm). Figure adopted from A. Gadeyne.

		MaxQuant						NSAF					
		T0 versus CTset T0 (AUR1/2)			T10 versus CTset T10 (PSB-D)			T0 versus large background			T10 versus large background		
Accession	Protein name	CCS52A1	CCS52A2	CCS52B	CCS52A1	CCS52A2	CCS52B	CCS52A1	CCS52A2	CCS52B	CCS52A1	CCS52A2	CCS52B
AT4G22910	CCS52A1	X			X			X			X		
AT4G11920	CCS52A2		X			X			X			X	
AT5G13840	CCS52B			X			X			X			X
AT5G05560	APC1	X	X	X	X	X	X	X	X	X	X	X	X
AT2G04660	APC2	X	X	X	X	X	X	X	X	X	X	X	X
AT3G16320	APC3a/CDC27a		X	X		X	X	X	X	X	X	X	X
AT2G20000	APC3b/HOBBIT	X	X	X	X	X	X	X	X	X	X	X	X
AT4G21530	APC4	X	X	X	X	X	X	X	X	X	X	X	X
AT1G06590	APC5	X	X	X	X	X	X	X	X	X	X	X	X
AT1G78770	APC6/Nomega	X		X	X	X	X	X	X	X	X	X	X
AT2G39090	APC7	X	X	X	X	X	X	X	X	X	X	X	X
AT3G48150	APC8	X	X	X	X	X	X	X	X	X	X	X	X
AT2G18290	APC10		X	X			X	X	X	X	X		X
Q8H1U3*	APC12	X	X	X	X	X	X	X	X	X	X	X	X
AT1G73177	APC13	X	X	X		X	X	X	X	X	X	X	X
AT5G63135	APC15			X				X	X	X	X	X	X
AT2G42260	UVI4/PYM			X			X		X	X			X
AT3G57860	OSD1/UVI4-Like/GIGAS	X	X	X	X	X	X	X	X	X	X	X	X
AT1G32310	SAMBA			X		X	X		X	X	X	X	X
AT1G53140	DRP5A | Dynamin related protein 5A	X	X	X	X	X	X	X	X	X	X	X	X
AT2G45700	sterile alpha motif (SAM) domain-containing protein	X		X	X		X	X	X	X	X	X	X
AT1G03780	TPX2 | targeting protein for XKLP2				X		X	X	X	X	X	X	X
AT1G34355	ATPS1, PS1			X			X	X	X	X	X	X	X
AT4G14310	Transducin/WD40 repeat-like superfamily protein		X	X			X	X	X	X	X	X	X
AT2G27970	CKS2							X	X	X	X	X	X
AT3G48750	CDKA;1						X		X	X			X
AT5G43080	CYCA3;1								X	X			
AT1G47230	CYCA3;4								X	X			
AT5G19330	ARIA | ARM repeat protein interacting with ABF2	X						X	X		X		
AT1G22730	MRF2							X		X			
AT4G01370	ATMPK4, MPK4 | MAP kinase 4		X	X			X		X	X			X
AT2G20480	unknown protein			X			X		X	X			X
AT4G32830	AUR1								X	X			
AT4G18800	RABA1d							X			X		
AT5G23900	Ribosomal protein L13e family protein										X		
AT5G46430	Ribosomal protein L32e										X		
AT4G05520	EHD2								X				
AT4G17650	Polyketide cyclase								X				
AT3G07090	PPPDE putative thiol peptidase family protein								X				
AT5G48810	ATB5-B											X	
AT5G03740	Histone deacetylase 2C											X	
AT1G31860	HISN2											X	
AT5G54500	FQR1											X	

AT5G11510	MYB3R-4, AtMYB3R4 | myb domain protein 3r-4			X			X		X		X
AT1G51690	ATB ALPHA, B ALPHA			X			X		X		X
AT3G58500	PP2A-4								X		X
AT5G60930	KIF4A ortholog								X		X
AT4G18600	WAVE5								X		X
AT2G25880	AUR2								X		X
AT3G14740	RING/FYVE/PHD zinc finger superfamily protein			X			X		X		X
AT3G09880	ATB~ BETA								X		X
AT5G23910	KIF2C ortholog								X		
AT4G27060	TORTIFOLIA 1								X		
AT4G17330	ATG2484-1, G2484-1 | G2484-1 protein			X					X		
AT5G25590	Protein of unknown function								X		
AT1G14510	AL7								X		
AT2G24200	LAP1										X
AT4G24690	NBR1								X		
AT2G19540	Transducin family protein / WD-40 repeat family protein	X		X	X	X					
AT4G25730	FtsJ-like methyltransferase family protein	X									
AT5G16070	TCP-1/cpn60 chaperonin family protein	X	X	X							
AT3G22660	rRNA processing protein-related	X									
AT3G10410	SCPL49, CPY | SERINE CARBOXYPEPTIDASE-LIKE 49			X							
AT1G05910	cell division cycle protein 48-related / CDC48-related			X							
AT1G24300	GYF domain-containing protein			X							
AT2G38250	Homeodomain-like superfamily protein			X							
AT1G07370	PCNA1, ATPCNA1						X				
AT2G42570	TBL39 | TRICHOME BIREFRINGENCE-LIKE 39						X				
AT3G50370	unknown protein						X				
AT5G01380	Homeodomain-like superfamily protein						X				
AT5G05780	RPN8A, AE3, ATHMOV34 | RP non-ATPase subunit 8A						X				

Supplementary table 1: Comparison between different filtering approaches for proteins identified by liquid chromatography-tandem MS (LC-MS/MS) after AP-MS on the three APC co-activators CCS52A1, CCS52A2 and CCS52B. Three AP-MS analyzes for every bait and every time point were executed. MaxQuant: Proteins were identified via MaxQuant after 4 hours propyzamide treatment (T0) and 10 min after removal of propyzamide (T10). Significant enriched proteins at T0 were identified relative to six AUR1 and AUR2 T0 pull-downs with a FDR = 0.0001, S0 = 1. Significant enriched proteins at T10 were identified relative to three PSB-D T10 pull-downs with a FDR = 0.001, S0 = 1. NSAF: Proteins significant enriched after filtering relative to a large background list are represented. *APC12 was retrieved with MaxQuant analysis relative to a control data set using the Araport11 database with APC12 (Q8H1U3) protein sequence from UniProt added.

Discussion and future perspectives.

Caroline Matthijs[1,2], Astrid Gadeyne[1,2], Jelle Van Leene[1,2], Geert De Jaeger[1,2]

Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium[1]

VIB Center for Plant Systems Biology, 9052 Ghent, Belgium[2]

Author contributions

CM wrote the chapter. AG, JVL and GDJ supervised and complemented the writing.

Introduction

To gain a deeper insight on how an organism functions, the complex interplay between the different organs, tissues, cells, and on a molecular level between proteins, genes, metabolites, etc., needs to be mapped. To do so, a range of tools exist to generate multiple '-omic' data sets, including genomic, phenomic, transcriptomic, proteomic and interactomic data. In this doctoral research we focused on two interactomic subdomains in plants, being protein-DNA interactions (PDIs) and protein-protein interactions (PPIs). For both interactions, an extensive list of techniques is available to reveal the interaction between two or multiple macromolecules under native or non-native conditions (reviewed in chapter 1 and 2). Different strategies can be applied, making use of reporter systems, fluorescence, bioluminescence, affinity purification or proximity labeling, each with their advantages and disadvantages. While the list of PPI and PDI analysis tools keeps expanding, there still remains unexplored territory. For plant PDIs a gene-centered method is not available yet, making the analysis of a specific plant genomic region dependent on ectopic Y1H screens or *in vitro* methods. On the other hand, studying transient PPIs by applying affinity purification is challenging as most of the time only the stable components are retrieved and the weak, transient interactions are lost. Both problems have been tackled differently by independent research groups, leading to suboptimal results. The gene-centered method proteomics of isolated chromatin segments (PICh) has been applied once in barley for the identification of interacting proteins on the centromeric chromatin (Zeng and Jiang, 2016). Different histone variants were retrieved, including the cenH3 variants, but previously characterized centromeric proteins were not detected. To identify weak transient PPIs, chemical cross-linking has been implemented in affinity-based methods (Rohila et al., 2004; Tagwerker et al., 2006; Stingl et al., 2008; Van Leene et al., 2019). The resulting covalent bonds between interacting proteins ensures that transient interactions are not lost during the necessary washing steps. However, cross-linking results in a higher number of false positives and additionally, it only reveals PPIs at the time point of cross-linking, limiting the amount of co-purified interacting partners.

In recent years, biotin-based proximity labeling has been introduced in the interactomic field and has led to the development of a new generation of PPI and PDI techniques (Roux, 2013; Rees et al., 2015; Schmidtmann et al., 2016; Myers et al., 2018). Proximity labeling relies on an enzyme, typically a biotin ligase or ascorbate peroxidase, that is capable of covalently labeling, e.g. biotinylation, proteins in the immediate vicinity. By fusing this enzyme to a target protein or by recruiting it to the genomic region of interest, interacting partners will be labeled and subsequently

purified by affinity purification and identified via mass spectrometry. Proximity labeling seems to be an ideal method to capture all interacting proteins, including stable, transient and dynamic interactions. However, proximity labeling is prone for false positive results due to endogenous biotinylated proteins and labeling of non-interacting proteins that are in the neighborhood of the target. As a consequence, analysis of appropriate negative controls is essential.

In this doctoral research the main focus was the development of an *in planta* gene-centered PDI tool by exploring three different strategies. First, we tested a multifunctional T-DNA construct making use of a tagged exogenous DNA binding protein and its corresponding DNA binding element for the specific isolation of different promoter sequences and its interacting partners (chapter 3). Later on, we combined this strategy with proximity labeling, allowing *in vivo* biotinylation and a stringent streptavidin-based purification of interacting proteins (chapter 3). As a third option we tested the applicability of CRISPR for the endogenous targeting and subsequent isolation of multicopy loci and their proteome (chapter 4). Furthermore, we focused on the further validation of an in-house developed pull-down technique (AP-MS) by identifying the weak, transient interactions between the anaphase promoting complex (APC) and its mitotic substrates (chapter 5). In addition, we tested if proximity labeling could generate similar or even superior results for the identification of APC substrates.

Exploring the plant chromatin landscape

Chromatin affinity purification and the disadvantage of chemical cross-linking.

In chapter 3 we tried to develop a new gene-centered in planta tool by adopting the modus operandi of ChAP-MS and iChIP in our *Arabidopsis* cell suspension culture. A T-DNA construct was created, consisting of the DNA sequence of interest (e.g. a promoter) flanked by the prokaryotic LacO sequence and a cassette allowing expression of GS^{rhino}-tagged prokaryotic LacI. Functionality of this T-DNA in our PSB-D culture was demonstrated by efficient accumulation of GS^{rhino}-LacI and binding of the ARF7 TF on the DR5v2 promoter. Pull-down of GS^{rhino}-LacI also resulted in an effective enrichment of the promoter sequence of interest, illustrating an efficient binding of the lac repressor on LacO in plant cells. However, subsequent label free quantitative MS analysis of the DR5v2, ETG1 and CycB1;2 promoters did not result in the specific enrichment of expected upstream regulators. A large amount of non-specific interactions was retrieved, which is a general disadvantage of one-step purifications. We identified formaldehyde cross-linking as a culprit for the aggravated

co-purification of non-specific interactions, which prevents us of filtering out the genuine protein interactions of a specific DNA locus.

Formaldehyde fixates all macromolecular interactions that are in close proximity (± 2 Å) (Hoffman et al., 2015). This includes fixation of transcription regulators on their corresponding DNA binding motif, co-regulators binding transcription factors, proximal enhancer regions that are recruited to a promoter region and chromatin associated RNA's. Surprisingly, formaldehyde is not able to fixate every protein-DNA interaction as it has been shown *in vitro* that purified LacI is not cross-linked by formaldehyde to LacO-containing DNA (Solomon and Varshavsky, 1985). In addition, *in vivo* cross-linking was unnecessary to retrieve the binding between LacI and LacO in chloroplasts via a ChIP experiment, which means that the strong affinity (K_d of ~10^{-13} M) between LacI and LacO can withstand affinity purification (Newell and Gray, 2010). However, we were unable to purify our promoter sequence of interest without cross-linking using the LacI-LacO interaction, while with cross-linking a significant enrichment was observed, which contradicts the *in vitro* observation by Solomon and Varshavsky (1985). Additionally, cross-linking was necessary to co-purify the promoter specific interacting proteins, sadly accompanied with the pull-down of a massif amount of non-specific interactions. Because of the dense concentration of macromolecules in the nucleus and the fixation between different macromolecule types, formaldehyde cross-linking has shown to result in the formation of higher order networks, thereby linking unrelated chromatin regions with each other, illustrated in figure 1 (Gavrilov et al., 2015). As a consequence, non-specific interactions are being co-purified during affinity purification of TFs or specific genomic regions.

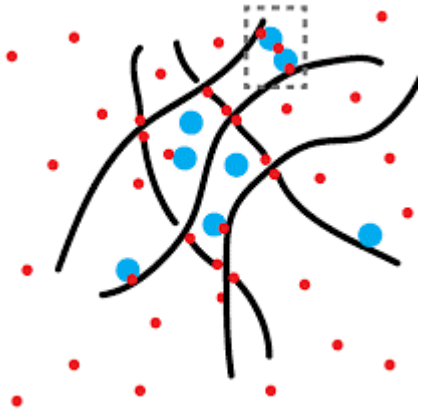


Figure 1: Potential effects of formaldehyde in mediating formation of higher order chromatin structures. The black wavy lines denote chromatin fibers, which may become a cross-linked meshwork in the presence of formaldehyde (red circles). The formation of these potentially confounding structures may or may not be mediated by physiologically relevant higher order interactions captured by crosslinking (dashed gray rectangle). Such a meshwork may define localized neighborhoods in the nucleus that trap proteins (cyan) that may or may not interact specifically with nearby DNA sequences in an unperturbed cell. Figure adopted from Hoffman et al. (2015).

Alternative protein-DNA cross-linkers exist, among which cis-diammine dichloro platinum II (cis-DDP or cisplatin) has been used to a limited extent for ChIP analyses (Chichiarelli et al., 2002; Cervoni et al., 2003; Chichiarelli et al., 2007). Cisplatin only forms protein-DNA complexes with a low reactivity towards histones (Pinto and Lippard, 1985) and the cross-linking can be reversed by the use of thiourea. This

alternative cross-linking with cisplatin could avoid the formation of higher order chromatin networks and thus result in a more efficient identification of bona fide direct interactors with MS. However, indirect interactions, like co-activators, would be lost, eventually leading to an incomplete understanding of the transcriptional regulation of a specific DNA locus. Another alternative for *in vivo* cross-linking is the application of ultraviolet (UV) light. UV-cross-linking induces covalent bonds between macromolecules that are in close contact with each other. Next to its common usage for the discovery of protein-RNA interactions (Sugimoto et al., 2012; Urdaneta and Beckmann, 2019), it also has been applied for the identification of protein-protein and protein-DNA interactions (Chodosh, 2001; Zhang et al., 2004; Itri et al., 2016), even in plant cells (Marondedze et al., 2016). An advantage of UV is the very short irradiation time that is needed to cross-link *in vivo* interactions, especially when an UV-laser is applied, a nano-, pico- or even femtosecond can be enough to generate covalent bonds between interacting macromolecules (Zhang et al., 2004; Itri et al., 2016). In comparison, formaldehyde cross-linking requires an average of 5 seconds to fixate interactions, and is as such unable to fix transient interactions (Schmiedeberg et al., 2009). It has been reported that TF DNA-binding can be very dynamic with a mode of action that is called ‘hit-and-run’ (Charoensawan et al., 2015; Doidy et al., 2016). This includes the transient binding of a target sequence, activating expression and subsequent dissociation from the activated target gene. The short residence time causes the need for a rapid freezing of the PDI and can be achieved via UV-cross-linking. Replacing formaldehyde cross-linking with UV-cross-linking in our ChAP-MS protocol, could avoid the formation of higher order networks and would result in a lower level of non-specific co-purified proteins while there is a higher chance to pull-down transient bona-fide interactions. This also applies to our CRISPR-ChAP-MS approach, but UV-cross-linking will not resolve the pitfall we encounter there (chapter 4). In contrast to ChAP-MS, purification of multi-copy loci with CRISPR-ChAP-MS was not accomplished. A kinetic study has reported that Cas9 dissociation from DNA is extremely slow (Raper et al., 2018), thus formaldehyde treatment would be sufficient to fixate the ternary complex of dCas9, guide RNA and DNA. We postulate that the lack of target sequences after pull-down via CRISPR-ChAP-MS is due to a low binding rate or even absence of dCas9 binding.

CRISPR-ChAP-MS: room for improvement.

For CRISPR-ChAP-MS (chapter 4) we used a human codon optimized *Streptococcus pyogenes* Cas9 (spCas9) which has been applied in different plant species like *Arabidopsis*, tobacco, poplar and rice (Feng et al., 2013; Mao et al., 2013; Nekrasov et al., 2013; Xie and Yang, 2013; Zhou et al., 2015). Although spCas9 is the most widely used CRISPR nuclease, it also has some limitations. It is prone for targeting non-specific DNA sequences, which we have noticed as well with the co-purification of off-target sequences for ribosomal DNA (rDNA). In addition, spCas9 targeting requires the presence of a specific short stretch of nucleotides, termed the protospacer adjacent motif (PAM) sequence (5'-NGG-3'), downstream of the target sequence. This limits the number of possible gRNAs to target a specific locus, and could even prevent gRNAs from being developed for certain target sites. Furthermore, the large size of spCas9 can hinder transformation efficiency, and in our *Arabidopsis* cell suspension culture we observed a significant amount of dCas9 degradation. To circumvent these drawbacks, alternative Cas9 variants can be used in gene-centered PDI methods, as has been reported recently in mammalian cells for the engineered DNA binding molecule-mediated chromatin immunoprecipitation (enChIP) approach (Fujita et al., 2018). Instead of using the spCas9, Fujita and co-workers used the *Staphylococcus aureus* Cas9 (saCas9), which has a smaller size (30 kDa smaller than spCas9) and a less stringent PAM sequence (5′-NNGRRT-3′ or 5′-NNGRR(N)-3′), allowing the development of a more flexible enChIP method. Another Cas9 variant that is interesting to incorporate in the CRISPR-ChAP-MS strategy is the *Streptococcus canis* Cas9 (scCas9), which has a less strict PAM sequence (5'-NNG-3') consequently increasing the amount of genomic targets (Chatterjee et al., 2018). Furthermore, Cpf1 (CRISPR from *Prevotella* and *Francisella 1*) can be applied when targeting AT-rich DNA loci (Zetsche et al., 2015). To avoid off-targets, engineered nucleases have been developed with improved specificity (Kleinstiver et al., 2016; Slaymaker et al., 2016; Chen et al., 2017). The possibilities to further optimize the CRISPR-ChAP-MS strategy are even rising with the development of small affinity tags with high binding efficiency (Liu et al., 2017; Fujita et al., 2018) and the recent advancements in proximity labeling (Schmidtmann et al., 2016; Myers et al., 2017; Gao et al., 2018). Because gene-specific protein-DNA interactions have a low abundancy and MS identification of proteins demands enough protein yield, multi-copy genes, especially telomeres, have been the most analyzed targets to obtain proof of concept. However, recent advancements in MS development creating more sensitive devices with a detection limit ranging between pico- and femtomole protein (Donnelly et al., 2019), and the ongoing application of CRISPR-based gene-centered methods, reports on the analysis of single loci are rising, proving that the CRISPR strategy will likely be the method of

choice in the future to analyze PDIs for a specific genomic locus (Campbell et al., 2018; Hamidian et al., 2018; Han et al., 2018).

When we would further optimize our plant-specific CRISPR-ChAP-MS approach, it would be wise to test the above mentioned adjustments and not only target multi-copy loci, like telomeres and rDNA which are in our hands not ideal targets, but also single loci in an active state allowing dCas9 binding. Especially for the latter, our *Arabidopsis* cell suspension culture (PSB-D) could be beneficial. Because of its high ploidy level (9C), targeting a single genomic locus in PSB-D can result in the purification of multiple PDI complexes per cell, increasing the amount of genuine interacting proteins and increasing the chance to identify them with MS. In addition, our cell culture is susceptible for different stimuli, e.g. sucrose starvation and repletion, aphidicolin and propyzamide treatment (chapter 5) (Menges and Murray, 2002; Van Leene et al., 2019), which can be used to activate expression of target genes and stimulating dCas9 binding. Supplementary table 1 displays some calculations to estimate the amount of starting material needed to obtain a sufficient protein yield (1 femtomole) for MS detection, comparing the ChAP-MS and CRISPR-ChAP-MS strategy for different targets.

Proximity labeling: a new generation of gene-centered tools on the horizon.

Recent years, proximity labeling has become more and more popular for the investigation of interaction landscapes (Roux, 2013; Rees et al., 2015) and it did not take long for its first applications in gene-centered PDI methods (Schmidtmann et al., 2016; Myers et al., 2017). By bringing a promiscuous biotin ligase (BirA*) or ascorbate peroxidase (APEX2) to the genomic region of interest, all proteins in proximity will be covalently labeled with biotin, allowing a stringent streptavidin-based purification. Due to the *in vivo* labeling, fixation of interactions via cross-linking and pull-down of the specific genomic locus becomes unnecessary, hence eliminating some pitfalls we encountered with (CRISPR-)ChAP-MS. Efforts have been made to develop more efficient proximity labeling strategies, among which a mutated version of the biotin ligase BirA with greater proximity labeling efficiency, called TurboID (Branon et al., 2018). This mutant BirA enzyme reduces the labeling time from more than 18 hours to 10 minutes using a lower amount of biotin. In addition, TurboID-based proximity labeling can be applied under normal plant growth conditions (Deepanksha Arora, 2019; Mair et al., 2019; Tae-Wuk Kim, 2019), paving the way for an *in planta* gene-centered method. In an initial experiment we tested the functionality of TurboID for the identification of interacting proteins of the synthetic DR5v2 promoter (chapter 3). However, similar to ChAP-MS, we stumble upon an extensive list of co-purified

proteins, yet with DR5v2 specifically enriched proteins, but from which many could be discarded as false positives (no nuclear localization) and common chromatin related proteins. Our data suggests a non-specific biotinylation of all types of cellular proteins, which we link to a high amount of free and highly active TurboID. A low signal-to-noise ratio using TurboID has also been reported for the analysis of the plasma membrane-associated octameric TPLATE complex (Deepanksha Arora, 2019). Further optimization is needed if we want to unravel the chromatin landscape in plant cells via proximity labeling. Limiting TurboID expression by putting it under control of an inducible promoter, will avoid the accumulation of endogenous biotin-based labeled proteins during the entire time-span of cell growth. Furthermore, TurboID should be directed to the nucleus to avoid labeling of non-nuclear proteins. Given the fact that the nucleus is a small subcellular compartment with a dense amount of proteins, special attention is required to overcome non-specific proximity labeling, especially when using the highly active TurboID. Limiting the biotinylation time-span would already diminish non-specific labeling, but including proper negative controls will remain a necessity to filter out non-specific labeled proteins, as has been previously reported for the identification of PPIs in plant nuclei with TurboID (Mair et al., 2019). Mair and co-workers suggest the incorporation of several negative controls including a wild type sample to correct for endogenous biotinylated proteins, a free nuclear localized TurboID sample to filter out non-specific labeled proteins and an untreated sample to eliminate proteins that are not labeled in response to biotin treatment (Mair et al., 2019). Alternatively, less active biotin ligases can be used, like BioID and BioID2 (Cronan, 2005; Kim et al., 2016), for which their functionality in plants has been reported (Lin et al., 2017; Conlan et al., 2018; Khan et al., 2018; Das et al., 2019; Deepanksha Arora, 2019). However, both function optimally at high temperatures (37°C for BioID and 50°C for BioID2 (Kim and Roux, 2016)), causing the need for long labelling times and very high biotin concentrations to gain some biotinylation under standard plant growth conditions. This low activity may be beneficial for lowering non-specific labeling in the dense nucleus, but will be inefficient for the detection of transient interactions including the dynamic interaction between a TF and its target sequence.

Further optimizing the TurboID proximity labeling strategy by combining it with CRISPR, in accordance with the GloPro and CasID methods in mammalian cells (Schmidtmann et al., 2016; Myers et al., 2017), could eventually lead to the generation of a gene-centered *in planta* method targeting genomic loci. Even then, account should be taken for the fact that the relative large TurboID enzyme (35 kDa) can hinder PDIs, and when analyzing macromolecular complexes, some interacting partners, hidden in the core of the complex, could be inaccessible for biotin labeling, leading to false negative results. In addition, TurboID only labels proteins, therefore missing out on

long-range DNA interactions and chromatin associated RNAs. To completely understand the regulation of a specific plant genomic locus, both gene-centered pull-down and proximity labeling methods are desirable. Although we were unsuccessful to identify interacting bona-fide proteins with our ChAP-MS strategy, we did not test for its applicability to isolate long-range DNA interactions or RNAs, as has been demonstrated for iChIP, enChIP and CAPTURE (Hoshino and Fujii, 2009; Fujita et al., 2013; Liu et al., 2018). While low abundant bona-fide interacting proteins are difficult to identify with MS, detection of low abundant DNA and RNA sequences is facilitated by the implementation of sequence amplification followed by next-generation sequencing or RNA-seq.

Single-cell approaches will further promote the efficiency of gene-centered methods, not only allowing identification of interacting DNA and RNA sequences put also proteins when combined with a nanopore-based protein sequencing approach (Swaminathan et al., 2018; Doerr, 2019). The latter makes use of the classic Edman degradation method to obtain sparse amino acid–sequence information for thousands to millions of proteins, followed by identification of the proteins by comparing amino acid-sequences against a reference proteome database (Swaminathan et al., 2018). While MS analysis requires a quantity of at least femtomole peptide for detection, this approach allows identification of a single peptide/protein and thus does not require large sample volumes or abundant protein levels.

Exploring low affinity interactions

When we unraveled the phosphorylation and interaction landscape of the plant target of rapamycin (TOR) kinase we observed that using our standard tandem affinity purification (TAP) protocol resulted mainly in the identification of stable interactions, being interactions between the three core complex subunits (TOR, LST8 and RAPTOR) and regulatory proteins (Van Leene et al., 2019). Additionally, only one substrate could be retrieved, being the translation initiation factor eIF2B-δ. Weak transient interactions, such as kinase-substrate interactions, are probably lost during the long double-step TAP purification. Therefore, we adjusted our TAP protocol to a one-step pull-down protocol (AP-MS) using home-made magnetic IgG beads which contain high binding capacity and are small in size, which results in less non-specific interacting proteins on the resin. These magnetic beads allow us to purify protein complexes using the strong affinity between ProtG and IgG. It also results in faster, integral washing steps, and shorter incubation times. As a result, we identified 8 proteins that interact with TOR and are phosphorylated in a TOR-dependent way, including three orthologues of known TOR substrates in yeast or mammalian cells.

Further validation via an *in vitro* kinase assay confirmed the presence of at least two other plant specific TOR substrates. To further validate the functionality of AP-MS for the identification of weak transient interactions, we analyzed the interactome of the APC E3-ligase complex during mitosis, using the three CCS52 co-activators as bait proteins. This as well revealed the interaction with orthologues of known substrates, and provided us with a list of new putative APC^{CCS52} targets.

Although both studies identified potential new substrates, known substrates like S6K for TOR and CycB1;1 for APC, were not retrieved. The necessary washing steps, the practical limitations (e.g. incomplete synchronized cell culture and incubation times) and the higher amount of non-specific interactions commonly pulled-down with one-step purifications, still causes the loss of some weak transient interactions or interferes with detection of low abundant proteins. To ensure retention of weak interactions during the pull-down, it would be worthwhile to apply cross-linking. Especially when a specific time point in development is targeted, as we have done for the CCS52 co-activators, cross-linking will fixate all interactions at that time point. Additionally, for transient interactions UV cross-linking would be favorable because of its extremely rapid reaction (nanoseconds or less) leading to covalent bonds between macromolecules that are in close contact. However, this could lead to an increase in non-specific interactions making it necessary to include proper negative controls to filter out the bona-fide interactions. Furthermore, more stringent washing steps by increasing salt or non-ionic detergent concentrations may reduce the amount of proteins that non-specifically bind with the affinity resin. By applying cross-linking, rigorous washing conditions can be applied without losing weak transient interactions. On the other hand, switching from ProtG-IgG to biotin-streptavidin pull-down, thereby making advantage of the strong non-covalent interaction between them, would ensure the retention of the bait protein on the resin during these stringent washing steps. In that respect, the biotin-based proximity labeling approach seems to be a good alternative for PPI analysis. Especially when using TurboID, which reduces the labeling time dramatically (Branon et al., 2018), making it beneficial for analyzing PPI dynamics, weak transient interactions and proteins that turn over rapidly *in vivo*.

To validate our CCS52B AP-MS data and to further expand the interactome, we applied TurboID-based proximity labeling on cells enriched for mitosis. Furthermore, we applied the 26S proteasome inhibitor MG132 to circumvent degradation of ubiquitinated substrates, as previously described (Coyaud et al., 2015). Although we identified APC subunits, regulators and the putative new interactor PS1, we could not further validate our AP-MS data. We identified the APC substrate CycB1;1, however this interaction was absent with MG132. We postulate that the absence of APC

substrates may be due to steric hindrance during CCS52B association on the APC complex, caused by the large TurboID (35 kDa) ligase, which is significantly larger than the GS^{rhino} tag (21 kDa). APC co-activators have two distinct protein domains, a C-terminal IR tail and an N-terminal C-box, both being essential for association with the APC complex (figure 2) (Matyskiela and Morgan, 2009; Yamano, 2019). Terminal fusion with TurboID, can interfere with the association or with the formation of all subunits to a functional APC complex. To prevent interference with the formation of the APC complex and co-activator association, it would be more ideal to position TurboID in a loop region relatively far away from the C-box, IR-tail and the substrate binding domain. Furthermore, steric hindrance could be reduced using the smaller miniTurboID (28 kDa), which does not contain the N-terminal DNA binding domain (Branon et al., 2018). Although miniTurboID is less stable than TurboID (Branon et al., 2018), its functionality in plants has been reported (Deepanksha Arora, 2019). In general, compared to AP-MS, proximity labeling is more dependent on the position of the tag in relation to the 3D structure of the complex, as only proteins in close proximity to the tag will be biotinylated. In addition to putative steric hindrance, the distance of the tag to the substrate could also explain the lack of substrates identified with TurboID. Alternatively, mitotic APC substrates could be identified by using the catalytic core subunits APC2 and APC11 as baits or APC10 that together with the co-activators determines substrate specificity. Despite biotinylation and identification of putative APC substrates, further validation (by e.g. *in vitro* ubiquitination assays) is needed to link candidate substrates with the different CDC20 and CCS52 co-activators. Also during AP-MS analysis of the TOR complex it became clear that analyzing different subunits as bait proteins is essential to increase substrates isolation. AP-MS on two of the three TOR subunits (LST8 and RAPTOR) revealed that 7 of the 8 putative substrates were retrieved with the regulatory subunit LST8, while with RAPTOR, which is responsible for recognizing substrates, only 5 of the 8 putative targets were identified.

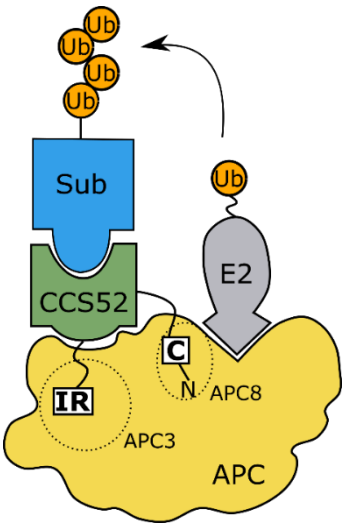


Figure 2: APC dependent ubiquitination of substrates. The isoleucine-arginine (IR) tail of the coactivator CCS52 binds to APC3 and the C-box interacts with APC8 for activation of the APC. Both IR tail binding and C-box binding ensure stable binding of the co-activator to the APC. The WD40 domain of the co-activator (green) is responsible for substrate degron recognition. Association of the ubiquitin (Ub) conjugating enzyme (E2) results in poly-ubiquitination of the substrate.

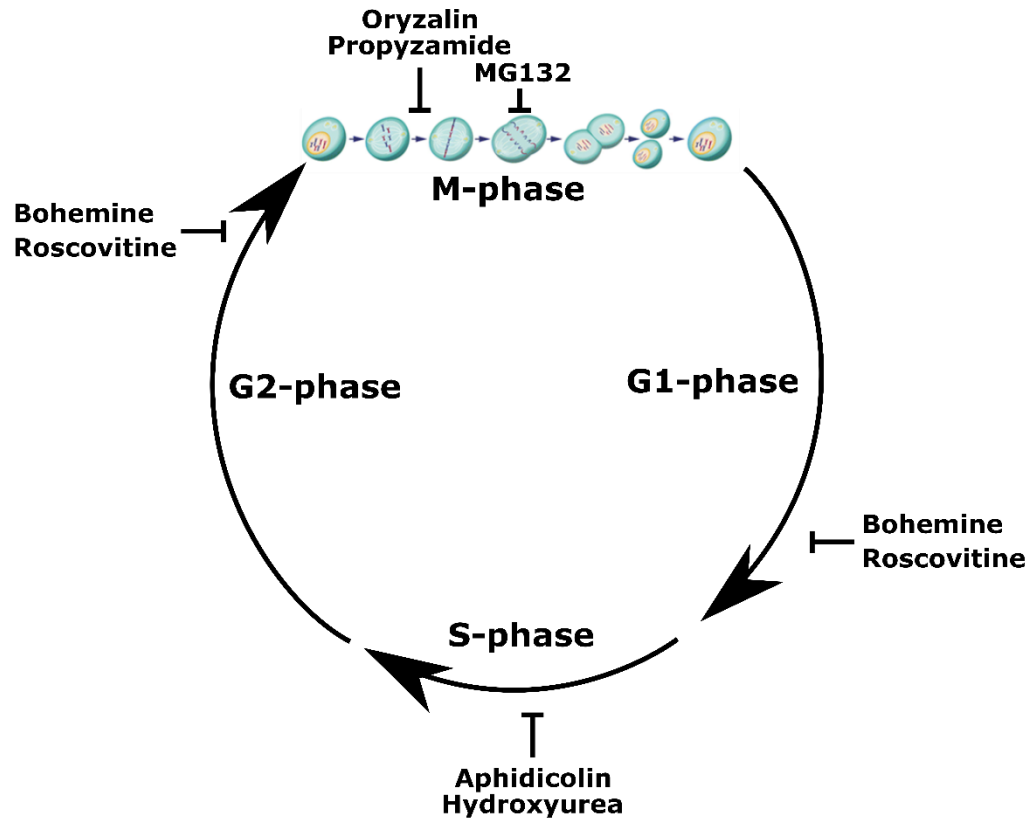
The list of PPI tools is extensive, but no single PPI assay is 100% efficient, including the most recently developed technologies. All methods have their advantages and disadvantages. Depending on the aim and the characteristics of the bait protein, one method will be more suitable than the other. A well-considered choice of method, bait protein and N- or C-terminal fusion protein must be made and a more elaborate interactome map will be gained by combining different complementary assays followed by *in vivo* validation.

A mitotic function for CCS52B

Our AP-MS data further provided evidence for a mitotic role of the APC co-activator CCS52B. Based on expression profiles and localization studies it was already postulated that CCS52B plays a main role during mitosis and fulfills the role of CDH1 in *Arabidopsis* (Menges et al., 2003; Yang et al., 2017). Our data further confirms this hypothesis by identifying interactions with different mitotic proteins including mitotic kinesins. Our CCS52B pull-down on propyzamide synchronized cell culture retrieved orthologues of known mitotic substrates in mammalian cells, including Tpx2 (Stewart and Fang, 2005) and the aurora kinases (Stewart and Fang, 2005; Floyd et al., 2008) which are required for the assembly and reorganization of the mitotic spindle during metaphase and anaphase. Also interactions with proteins that function at a later stage of mitosis were pulled-down, like the MPK4 which is essential for formation of the phragmoplast (Beck et al., 2011). Although we synchronized our cell culture with propyzamide to identify interactions at the metaphase-anaphase transition, proteins that play a role during the exit of mitosis were also found. It was shown before that 75 percent of the cells are not blocked at metaphase and CCS52B protein levels peak at cytokinesis (Yang et al., 2017). To obtain a higher synchronization level, a double-step synchronization is often used, where DNA synthesis at S-phase is blocked first using aphidicolin or hydroxyurea (figure 3A), followed by a washing step to allow cell cycle progression until metaphase for a second arrest using propyzamide or oryzalin (Planchais et al., 2000). Such a double-step synchronization strategy has been applied successfully in BY-2 cells with a synchronization level of almost 90% (Nagata et al., 1992). A similar approach utilizing our *Arabidopsis* cell suspension culture could result in a higher synchronization and may enrich for mitotic interactions with the APC co-activators. To further expand the interactome of CCS52B during mitosis, AP-MS should be applied at later time-points after mitotic synchronization. In addition, the mitotic role of the two CCS52 A-type co-activators should be investigated into more detail, as we also identified mitotic interactors for these proteins. Furthermore, AP-MS on the two functional CDC20 co-activators will reveal APC substrates at the early

stages of mitosis and could confirm the *in vivo* interaction with the recently identified *Arabidopsis* homologs for Securin, called PATRONUS 1 and PATRONUS 2 (PANS1 and PANS2) (Cromer et al., 2019). To target the early phases of mitosis, synchronization can be obtained by blocking the CDK activity with roscovitine or bohemine, resulting in cell cycle arrest at both G1/S and G2/M transitions (figure 3) (Planchais et al., 2000).

A



B

Name	Target	Block	Reversibility
Aphidicolin	DNA polymerases	S-phase	+
Hydroxyurea	ribonucleotide reductase	S-phase	+/-
Oryzalin	microtubule depolymerisation	metaphase	+/-
Propyzamide	microtubule depolymerisation	metaphase	+/-
Bohemine	CDK activity	G1/S and G2/M	+
Roscovitine	CDK activity	G1/S and G2/M	+
MG132	proteasome	metaphase/anaphase	-

Figure 3: Different chemicals arrest cell cycle progression at different phases. A) Simplified scheme of the cell cycle and the action time points for different blocking agents. **B)** The specificity of action of different chemical agents.

In accordance with the identification of TOR substrates through the integration of interactomics with phosphoproteomics (Van Leene et al., 2019), combining the APC interactome with a cell cycle based ubiquitination profile would bring more clarity on potential APC substrates. Proteome-wide ubiquitination studies are available (Maor et al., 2007; Saracco et al., 2009; Kim et al., 2013; Walton et al., 2016; Aguilar-Hernandez et al., 2017; Willems et al., 2019), but suffer from the low stability and time-specific presence of ubiquitinated proteins and the high activity of deubiquitylating enzymes (DUBs) in crude plant extracts. To identify targets of the APC complex, new ubiquitination mapping technologies (e.g. COFRADIC) can be used to differentially analyze wild-type and mutant CCS52 plants or mitotic synchronized cell cultures (Walton et al., 2016). Additionally, trypsin digested ubiquitinated peptides derived from proteins co-purified during AP-MS or proximity labeling studies can be specifically enriched by using an anti-diglycyl lysine (K-ε-GG) antibody (Xu and Jaffrey, 2013). The latter would also give us more insights in which co-purified proteins are substrates and which are regulatory proteins of the APC complex.

Taken together, our AP-MS data has laid a foundation for further research on the APC complex during mitosis in plants, especially for the not well characterized CCS52B co-activator. At the moment, putative substrates pulled-down with AP-MS (RabA1d, EHD2, MPK4, NBR1 and AUR2) are being validated by performing *in vitro* ubiquitination assays and cell free degradation studies, and their stabilization is being analyzed *in vivo* in wild-type and mutant ccs52 *Arabidopsis* lines, based on previous studies characterizing APC substrates (Boudolf et al., 2009; Lin et al., 2012). Furthermore, characterization of the putative degrons by mutation analysis should be examined. Next to the co-purification of known and new putative substrates, we also pulled-down some known and new putative regulators, including subunits of a protein phosphatase type 2A (PP2A) complex for which evidence has been found in mammalian cells and yeast that it regulates the phosphorylation status of the APC co-activators (Hein et al., 2017; Kataria and Yamano, 2019). *In vitro* assays following the phosphorylation status of the co-activators with increasing amounts of PP2A and mutant analysis should reveal the regulatory role of these PP2A subunits relative to the co-activators.

References

Aguilar-Hernandez V, Kim DY, Stankey RJ, Scalf M, Smith LM, Vierstra RD (2017) Mass Spectrometric Analyses Reveal a Central Role for Ubiquitylation in Remodeling the Arabidopsis Proteome during Photomorphogenesis. Molecular Plant **10:** 846-865

Beck M, Komis G, Ziemann A, Menzel D, Samaj J (2011) Mitogen-activated protein kinase 4 is involved in the regulation of mitotic and cytokinetic microtubule transitions in Arabidopsis thaliana. New Phytol **189:** 1069-1083

Boudolf V, Lammens T, Boruc J, Van Leene J, Van Den Daele H, Maes S, Van Isterdael G, Russinova E, Kondorosi E, Witters E, De Jaeger G, Inze D, De Veylder L (2009) CDKB1;1 forms a functional complex with CYCA2;3 to suppress endocycle onset. Plant Physiol **150:** 1482-1493

Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol **36:** 880-887

Campbell AE, Shadle SC, Jagannathan S, Lim JW, Resnick R, Tawil R, van der Maarel SM, Tapscott SJ (2018) NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife **7**

Cervoni L, Pietrangeli P, Chichiarelli S, Altieri F, Egistelli L, Turano C, Lascu I, Giartosio A (2003) In vivo cross-linking of nm23/nucleoside diphosphate kinase to the PDGF-A gene promoter. Mol Biol Rep **30:** 33-40

Charoensawan V, Martinho C, Wigge PA (2015) "Hit-and-run": Transcription factors get caught in the act. Bioessays **37:** 748-754

Chatterjee P, Jakimo N, Jacobson JM (2018) Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci Adv **4:** eaau0766

Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature **550:** 407-410

Chichiarelli S, Coppari S, Turano C, Eufemi M, Altieri F, Ferraro A (2002) Immunoprecipitation of DNA-protein complexes cross-linked by cis-diamminedichloroplatinum. Anal Biochem **302:** 224-229

Chichiarelli S, Ferraro A, Altieri F, Eufemi M, Coppari S, Grillo C, Arcangeli V, Turano C (2007) The stress protein ERp57/GRP58 binds specific DNA sequences in HeLa cells. J Cell Physiol **210:** 343-351

Chodosh LA (2001) UV crosslinking of proteins to nucleic acids. Curr Protoc Mol Biol **Chapter 12:** Unit 12 15

Conlan B, Stoll T, Gorman JJ, Saur I, Rathjen JP (2018) Development of a Rapid in planta BioID System as a Probe for Plasma Membrane-Associated Immunity Proteins. Front Plant Sci **9:** 1882

Coyaud E, Mis M, Laurent EM, Dunham WH, Couzens AL, Robitaille M, Gingras AC, Angers S, Raught B (2015) BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates. Mol Cell Proteomics **14:** 1781-1795

Cromer L, Jolivet S, Singh DK, Berthier F, De Winne N, De Jaeger G, Komaki S, Prusicki MA, Schnittger A, Guerois R, Mercier R (2019) Patronus is the elusive plant securin, preventing chromosome separation by antagonizing separase. Proc Natl Acad Sci U S A **116:** 16018-16027

Cronan JE (2005) Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. J Nutr Biochem **16:** 416-418

Das PP, Macharia MW, Lin Q, Wong SM (2019) In planta proximity-dependent biotin identification (BioID) identifies a TMV replication co-chaperone NbSGT1 in the vicinity of 126kDa replicase. J Proteomics **204:** 103402

Deepanksha Arora NBA, Chen Liu, Petra Van Damme, Lam Dai Vu, Anna Tornkvist, Francis Impens, Dominique Eeckhout, Alain Goossens, Geert De Jaeger, Thomas Ott, Panagiotis Moschou, Daniel Van Damme (2019) Establishment of Proximity-dependent Biotinylation Approaches in Different Plant Model Systems. BioRxiv

Doerr A (2019) Single-cell proteomics. Nat Methods **16:** 20

Doidy J, Li Y, Neymotin B, Edwards MB, Varala K, Gresham D, Coruzzi GM (2016) "Hit-and-Run" transcription: de novo transcription initiated by a transient bZIP1 "hit" persists after the "run". BMC Genomics **17:** 92

Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, Lippens JL, Aluri KC, Sarin R, Chen B, Lantz C, Jung W, Johnson KR, Koller A, Wolff JJ, Campuzano IDG, Auclair JR, Ivanov AR, Whitelegge JP, Pasa-Tolic L, Chamot-Rooke J, Danis PO, Smith LM, Tsybin YO, Loo JA, Ge Y, Kelleher NL, Agar JN (2019) Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nat Methods **16:** 587-594

Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y, Zhu JK (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res **23:** 1229-1232

Floyd S, Pines J, Lindon C (2008) APC/C Cdh1 targets aurora kinase to control reorganization of the mitotic spindle at anaphase. Curr Biol **18:** 1649-1658

Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, Fujii H (2013) Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci Rep **3:** 3171

Fujita T, Yuno M, Fujii H (2018) enChIP systems using different CRISPR orthologues and epitope tags. BMC Res Notes **11:** 154

Gao XD, Tu LC, Mir A, Rodriguez T, Ding Y, Leszyk J, Dekker J, Shaffer SA, Zhu LJ, Wolfe SA, Sontheimer EJ (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods

Gavrilov A, Razin SV, Cavalli G (2015) In vivo formaldehyde cross-linking: it is time for black box analysis. Brief Funct Genomics **14:** 163-165

Hamidian A, Vaapil M, von Stedingk K, Fujita T, Persson CU, Eriksson P, Veerla S, De Preter K, Speleman F, Fujii H, Pahlman S, Mohlin S (2018) Promoter-associated proteins of EPAS1 identified by enChIP-MS - A putative role of HDX as a negative regulator. Biochem Biophys Res Commun **499:** 291-298

Han B, Zhou B, Qu Y, Gao B, Xu Y, Chung S, Tanaka H, Yang W, Giuliano AE, Cui X (2018) FOXC1-induced non-canonical WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer. Oncogene **37:** 1399-1408

Hein JB, Hertz EPT, Garvanska DH, Kruse T, Nilsson J (2017) Distinct kinetics of serine and threonine dephosphorylation are essential for mitosis. Nat Cell Biol **19:** 1433-1440

Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem **290:** 26404-26411

Hoshino A, Fujii H (2009) Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. J Biosci Bioeng **108:** 446-449

Itri F, Monti DM, Della Ventura B, Vinciguerra R, Chino M, Gesuele F, Lombardi A, Velotta R, Altucci C, Birolo L, Piccoli R, Arciello A (2016) Femtosecond UV-laser pulses to unveil protein-protein interactions in living cells. Cell Mol Life Sci **73:** 637-648

Kataria M, Yamano H (2019) Interplay between Phosphatases and the Anaphase-Promoting Complex/Cyclosome in Mitosis. Cells **8**

Khan M, Youn JY, Gingras AC, Subramaniam R, Desveaux D (2018) In planta proximity dependent biotin identification (BioID). Sci Rep **8:** 9212

Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Roux KJ (2016) An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell **27:** 1188-1196

Kim DI, Roux KJ (2016) Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. Trends Cell Biol **26:** 804-817

Kim DY, Scalf M, Smith LM, Vierstra RD (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. Plant Cell **25:** 1523-1540

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature **529:** 490-495

Lin Q, Wang D, Dong H, Gu S, Cheng Z, Gong J, Qin R, Jiang L, Li G, Wang JL, Wu F, Guo X, Zhang X, Lei C, Wang H, Wan J (2012) Rice APC/C(TE) controls tillering by mediating the degradation of MONOCULM 1. Nat Commun **3:** 752

Lin Q, Zhou Z, Luo W, Fang M, Li M, Li H (2017) Screening of Proximal and Interacting Proteins in Rice Protoplasts by Proximity-Dependent Biotinylation. Front Plant Sci **8:** 749

Liu X, Zhang Y, Chen Y, Li M, Shao Z, Zhang MQ, Xu J (2018) CAPTURE: In Situ Analysis of Chromatin Composition of Endogenous Genomic Loci by Biotinylated dCas9. Curr Protoc Mol Biol**:** e64

Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, Cao H, Ni M, Liu Y, Gu Z, Dickerson KE, Xie S, Hon GC, Xuan Z, Zhang MQ, Shao Z, Xu J (2017) In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell **170:** 1028-1043 e1019

Mair A, Xu SL, Branon TC, Ting AY, Bergmann DC (2019) Proximity labeling of protein complexes and cell type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife **8**

Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu JK (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant **6:** 2008-2011

Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics **6:** 601-610

Marondedze C, Thomas L, Serrano NL, Lilley KS, Gehring C (2016) The RNA-binding protein repertoire of Arabidopsis thaliana. Sci Rep **6:** 29766

Matyskiela ME, Morgan DO (2009) Analysis of activator-binding sites on the APC/C supports a cooperative substrate-binding mechanism. Mol Cell **34:** 68-80

Menges M, Hennig L, Gruissem W, Murray JA (2003) Genome-wide gene expression in an Arabidopsis cell suspension. Plant Mol Biol **53:** 423-442

Menges M, Murray JA (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J **30:** 203-212

Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat Methods

Myers SA, Wright J, Zhang F, Carr SA (2017) CRISPR/Cas9-APEX-mediated proximity labeling enables discovery of proteins associated with a predefined genomic locus. Molecular & Cellular Proteomics **16:** S63-S63

Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco by-2 Cell-Line as the Hela-Cell in the Cell Biology of Higher-Plants. International Review of Cytology-a Survey of Cell Biology **132:** 1-30

Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol **31:** 691-693

Newell CA, Gray JC (2010) Binding of lac repressor-GFP fusion protein to lac operator sites inserted in the tobacco chloroplast genome examined by chromatin immunoprecipitation. Nucleic Acids Res **38:** e145

Pinto AL, Lippard SJ (1985) Binding of the antitumor drug cis-diamminedichloroplatinum(II) (cisplatin) to DNA. Biochim Biophys Acta **780:** 167-180

Planchais S, Glab N, Inze D, Bergounioux C (2000) Chemical inhibitors: a tool for plant cell cycle studies. FEBS Lett **476:** 78-83

Rabanal FA, Nizhynska V, Mandakova T, Novikova PY, Lysak MA, Mott R, Nordborg M (2017) Unstable Inheritance of 45S rRNA Genes in Arabidopsis thaliana. G3 (Bethesda) **7:** 1201-1209

Raper AT, Stephenson AA, Suo Z (2018) Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. J Am Chem Soc **140:** 2971-2984

Rees JS, Li XW, Perrett S, Lilley KS, Jackson AP (2015) Protein Neighbors and Proximity Proteomics. Mol Cell Proteomics **14:** 2848-2856

Rohila JS, Chen M, Cerny R, Fromm ME (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J **38:** 172-181

Roux KJ (2013) Marked by association: techniques for proximity-dependent labeling of proteins in eukaryotic cells. Cell Mol Life Sci **70:** 3657-3664

Saracco SA, Hansson M, Scalf M, Walker JM, Smith LM, Vierstra RD (2009) Tandem affinity purification and mass spectrometric analysis of ubiquitylated proteins in Arabidopsis. Plant J **59:** 344-358

Schmidtmann E, Anton T, Rombaut P, Herzog F, Leonhardt H (2016) Determination of local chromatin composition by CasID. Nucleus **7:** 476-484

Schmiedeberg L, Skene P, Deaton A, Bird A (2009) A temporal threshold for formaldehyde crosslinking and fixation. PLoS One **4:** e4636

Simon L, Rabanal FA, Dubos T, Oliver C, Lauber D, Poulet A, Vogt A, Mandlbauer A, Le Goff S, Sommer A, Duborjal H, Tatout C, Probst AV (2018) Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in Arabidopsis thaliana. Nucleic Acids Res **46:** 3019-3033

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. Science **351:** 84-88

Solomon MJ, Varshavsky A (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A **82:** 6470-6474

Stewart S, Fang G (2005) Anaphase-promoting complex/cyclosome controls the stability of TPX2 during mitotic exit. Mol Cell Biol **25:** 10516-10527

Stewart S, Fang G (2005) Destruction box-dependent degradation of aurora B is mediated by the anaphase-promoting complex/cyclosome and Cdh1. Cancer Res **65:** 8730-8735

Stingl K, Schauer K, Ecobichon C, Labigne A, Lenormand P, Rousselle JC, Namane A, de Reuse H (2008) In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification. Mol Cell Proteomics **7:** 2429-2441

Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M, Ule J (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol **13:** R67

Swaminathan J, Boulgakov AA, Hernandez ET, Bardo AM, Bachman JL, Marotta J, Johnson AM, Anslyn EV, Marcotte EM (2018) Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nat Biotechnol

Tae-Wuk Kim CHP, Chuan-Chih Hsu, Jia-Ying Zhu, Yuchun Hsiao, Tess Branon, Shou-Ling Xu, Alice Y Ting, Zhi-Yong Wang (2019) Application of TurboID-mediated proximity labeling for mapping a GSK3 kinase signaling network in Arabidopsis. bioRxiv

Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol Cell Proteomics **5:** 737-748

Urdaneta EC, Beckmann BM (2019) Fast and unbiased purification of RNA-protein complexes after UV cross-linking. Methods

Van Leene J, Han C, Gadeyne A, Eeckhout D, Matthijs C, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Van de Cotte B, Stes E, Van Bel M, Storme V, Impens F, Gevaert K, Vandepoele K, De Smet I, De Jaeger G (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nat Plants **5:** 316-327

Walton A, Stes E, Cybulski N, Van Bel M, Inigo S, Durand AN, Timmerman E, Heyman J, Pauwels L, De Veylder L, Goossens A, De Smet I, Coppens F, Goormachtig S, Gevaert K (2016) It's Time for Some "Site"-Seeing: Novel Tools to Monitor the Ubiquitin Landscape in Arabidopsis thaliana. Plant Cell **28:** 6-16

Willems P, Horne A, Van Parys T, Goormachtig S, De Smet I, Botzki A, Van Breusegem F, Gevaert K (2019) The Plant PTM Viewer, a central resource for exploring plant protein modifications. Plant Journal **99:** 752-762

Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol Plant **6:** 1975-1983

Xu G, Jaffrey SR (2013) Proteomic identification of protein ubiquitination events. Biotechnol Genet Eng Rev **29:** 73-109

Yamano H (2019) APC/C: current understanding and future perspectives. F1000Res **8**

Yang W, Wightman R, Meyerowitz EM (2017) Cell Cycle Control by Nuclear Sequestration of CDC20 and CDH1 mRNA in Plant Stem Cells. Mol Cell **68:** 1108-1119 e1103

Zeng Z, Jiang J (2016) Isolation and Proteomics Analysis of Barley Centromeric Chromatin Using PICh. J Proteome Res **15:** 1875-1882

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell **163:** 759-771

Zhang L, Zhang K, Prandl R, Schoffl F (2004) Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. Biochem Biophys Res Commun **322:** 705-711

Zhou X, Jacobs TB, Xue LJ, Harding SA, Tsai CJ (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate:CoA ligase specificity and redundancy. New Phytol **208:** 298-301

Supplementary data

Supplementary table 1: Estimation of the amount of starting material needed to obtain 1 femtomole of protein, allowing detection with mass spectrometry. ChAP-MS and CRISPR-ChAP-MS are compared for different targets (single and multicopy loci) assuming that the efficiency of these strategies is 10%. In case of ChAP-MS we assume that one T-DNA construct is transformed per cell and that the bait promoter interacts with the TF in every cell. Because of the 9C ploidy level in our cell culture, CRISPR-ChAP-MS can target a single locus nine times in one cell. Multicopy loci 5S rDNA and 45S rDNA have a copy number of +/- 2000 (Simon et al., 2018)[1] and +/- 800 (Rabanal et al., 2017)[2] respectively in Arabidopsis thaliana. The ploidy level increases this copy number 9 times. Because of the dosage control (many rDNA is situated in heterochromatin) we assume that only 25% of the rDNA is available for pull-down.

Method	Target	No. Repeats /cell	Interactions with target	No. cells
ChAP-MS	bait promoter	1	100%	6×10^9
CRISPR-ChAP-MS	single locus	9	100%	7×10^8
CRISPR-ChAP-MS	5S rDNA	18000[1]	25%	13×10^5
CRISPR-ChAP-MS	45S rDNA	7200[2]	25%	3×10^6

CV

Caroline Matthijs

Biotechnologist

27/02/1990 Zottegem

Tweekerkenstraat 146
9620 Zottegem

0494 58 84 63

09 256 18 78

c_matthijs@hotmail.com

in/caroline-matthijs

profile/Caroline_Matthijs

Professional experiences

Predoctoral fellow **Oct 2015 – Dec 2019**

VIB-UGent Center for Plant Systems Biology
Development and validation of new interactomic tools in plants.
- Experimental design and implementation
- Guidance of 2 students and co-workers
- Presenting results at conferences, group meetings and multidisciplinary meetings
- Co-authorship in Nature Plants

Junior researcher **Jan 2014 – Oct 2015**

VIB-UGent Center for Plant Systems Biology
Development of a new gene-centered analysis tool for protein-DNA interactions in Arabidopsis.
- Experimental design and implementation
- Guidance of 1 thesis students and co-workers
- Presenting results at group meetings and multidisciplinary meetings
- Writing and defending grant proposals.

Summer job **Aug 2012 – Sept 2012**

BASF CropDesign
Rice transformation and regeneration

Education

Master of science in Biochemistry and Biotechnology **UGent** **2013**

Bachelor in Biochemistry and Biotechnology **UGent** **2011**

Conferences

Plant Biology – Montreal, Canada **2018**
Poster presentation: ChAP-MS: Exploring the plant chromatin proteome.
Caroline Matthijs, Jelle Van Leene, Nancy De Winne, Eveline Van De Slijke, Geert Persiau, Dominique Eeckhout, Geert De Jaeger

Genome Engineering – Ghent, Belgium **2016**
Poster presentation: Development of a CRISPR-based ChAP-MS approach for identification of protein-DNA interactions in plants.
Caroline Matthijs, J. Van Leene, N. De Winne, E. Van De Slijke, D. Eeckhout, G. De Jaeger

European Plant Science Retreat – Amsterdam, the Netherlands **2016**
Poster presentation: Development of a new gene-centered analysis tool for protein–DNA interactions in A. thaliana.
Caroline Matthijs, Jelle Van Leene, Geert De Jaeger

Personal information

I'm a fast-learning, rational thinking and feedback driven researcher with hands-on experience in molecular analysis of plant protein interactions. During my years as a researcher at VIB-UGent I learned to solve problems and create new analysis tools both independently and with cross functional collaborations. I love to work on innovative biotech tools and like to be involved in their application for society.

Languages

Dutch	Native
English	Fluent
French	Sufficient
German	Basic

Hobby

Volleyball

Scientific publication

Van Leene, J., Han, C., Gadeyne, A., Eeckhout, D., **Matthijs, C.**, Cannoot, B., De Winne, N., Persiau, G., Van De Slijke, E., Van De Cotte, B., Stes, E., Van Bel, M., Storme, V., Impe, F., Gevaert, K., Vandepoele, K., De Smet, I., De Jaeger, G. (2019) Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. Nature Plants 5, 316-327

Workshops

Career Guidance Course – VIB training	2019
Creative thinking – UGent	2019
Initiation GIMP and Inkscape – VIB training	2017
CRISPR-based Genome Engineering – VIB training	2016
Summer school “Let’s talk science!” – UGent	2016
Precision Genome Engineering – VIB training	2016
Research Ethics - VIB training	2015

Guidance

Master thesis	Evi Ceulemans	2016

Title: Identificatie van transcriptiefactoren betrokken in osmotische stress via ChAP-MS in planten.

Master 1 project	Thibaut Sanders	2017

Title: Preparing for the search of key interactors in the SnRK1 pathway.

Erasmus+ traineeship	Luis Miguel Luengo Cerron	2018
Assistance practical courses	Plant physiology	2015 / 2016

Scientific skills

Interactomics
Tandem affinity purification with mass spectrometry (TAP-MS)
Immunoprecipitation (IP)
TurboID-based proximity labeling
Chromatin immunoprecipitation (ChIP)

Molecular Biology
Molecular cloning
Protein, DNA & RNA extraction
qPCR & real-time PCR
SDS-PAGE & Immunoblotting
Transient expression assay in N. tabacum
Heterologous expression and purification of proteins in bacteria

Plant genetics
Agrobacterium-mediated transformation of A. thaliana (plants & cell culture)
CRISPR
Genotyping

Analytical instrumentation
Confocal microscopy

Others
Writing and presenting scientific reports (Word, PowerPoint, Excel)
CLC Main Workbench
EndNote
Inkscape