GHENT UNIVERSITY – FACULTY OF SCIENCES
DEPARTMENT OF PLANT BIOTECHNOLOGY AND BIOINFORMATICS
VIB CENTRE FOR PLANT SYSTEMS BIOLOGY

CHARACTERISATION OF GROWTH REGULATORY PROTEIN COMPLEXES IN PLANTS THROUGH MULTIFUNCTIONAL TAP-TECHNOLOGY

Nienke Besbrugge

Academic year 2016-2017
Promoter: Prof. Dr. Geert De Jaeger

Thesis submitted as partial fullfillment of the requirements for obtaining the degree of Doctor of Philosophy in Science: Biochemistry and Biotechnology

'Look deep into nature, and then you will understand everything better.'

Albert Einstein

'Technology is destructive only in the hands of people who do not realize that they are one and the same process as the universe.'

Alan Watts

Dankwoord

Waauw, het dankwoord. Met in mijn hoofd op de achtergrond "This is the End" van the Doors. Het is zo ver, en ik had er mij eerlijk gezegd iets anders van voorgesteld. Maar dat neemt uiteraard niet weg dat ik een heel pak mensen te bedanken heb, want alleen had ik hier nooit kunnen staan (en ook niet willen staan).

Laten we beginnen bij het begin: Ik heb hier nu 5 jaar mogen rondlopen in de **functional interactomics** groep. Allemaal natuurlijk dankzij **Geert De Jaeger**, die na mijn thesis mijn wervelwind capaciteiten en wetenschappelijk denken blijkbaar goed genoeg vond om samen een technologisch project te schrijven. Een project dat mij aansprak vanaf dag 1 omdat dingen efficient maken mijn hobby is. En verkopen dat ik mijn technologie kan! Bedankt, chef, om samen die beurs bij het IWT binnen te halen, om ondanks de dalen toch de stijgende lijnen te blijven zien, om samen te kunnen nadenken over onze data en om zo een aangename en begripvolle promoter te zijn! Gelukkig heb je dit niet alleen hoeven te doen, en daarmee kom ik direct bij **Jelle Van Leene** en **Hilde Nelissen** uit. Dank jullie allebei! Wat zal het voor jullie rustig worden nu ik wegga! Bedankt voor de ontelbare antwoorden, de leuke avonden na het werk en al jullie tijd en vriendschap. Ik heb zoveel mogen leren in de afgelopen 5 jaar, ik zou er nog een boek over kunnen schrijven! Voorts wil ik ook graag de mensen bedanken die mij dag in dag uit hebben weten rondlopen in mijn witte jas (of liever, mijn onzichtbare jas zoals ik hem graag noem. Dat kleur staat mij nu eenmaal beter), en die tevens een groot deel van de technische support hebben geleverd voor mijn doctoraat: **Geert Persiau**, bedankt om mij in wondere wereld van TAP in te leiden, menig protocol heb ik aangepast met tips en tricks van jouw werk; **Nancy De Winne** en **Eveline Van De Slijke**, bedankt voor het voorzien van al het biologische materiaal en alle kleine beetjes hulp en uitleg die ik van jullie altijd kon verwachten. **Dominique Eeckhout**, je denkt waarschijnlijk dat mijn Alzheimer-op-jonge-leeftijd al ver gevorderd is, want telkens opnieuw moest je mij herinneren aan de stappen in de data-analyse, die niemand meer beheerst dan jij. Dank voor je engelengeduld, de dataverwerking (ontzettend veel dank hiervoor!), en je luisterend oor. **Bernard Cannoot,** 't is niet omdat je niet meer in de FI zit dat je er geen deel van blijft uitmaken, bedankt om mij daar zolang gezelschap te houden in ons hoekske, voor alle culturen en plantenwerkjes, en voor het verdragen (al dan niet met gemopper) van mijn brede muzieksmaak. **Michiel Bontinck** en **Caroline Matthijs**, the next generation doctors, misschien viel dit voor jullie niet zo op, maar jullie aanwinst aan de groep was voor mij een blije verandering, want tot dan had ik nog geen partner doctors in crime, en alleen is maar alleen! Ik hoop dat jullie binnenkort ook kunnen beginnen afronden, en dat de vervolgcarrieres vruchtvol zullen zijn. Of op z'n minst goed betaald. **Astrid Gadeyne**, het was een blij weerzien na al die Jaren! Ik weet nog dat ik in jouw dankwoord stond als groentje die haar IWT had aangevraagd, wel kijk, nu sta jij in het mijne als die lieve post-doc die terugkwam met kalmte, raad, koffiebabbels, en groepverenigende eetactiviteiten. Alle success met je tarweproject, meid! **Lana Hellebaut** en **Dora Bordoni**, thank you so much for the lovely diner nights, the listening ears, the laughter and the joy we shared. Hope your scientific projects will be successful and that our paths may cross again in the future (I mean, come on, they have to)! The same counts for **Clara Williams**, my knuffle waffle, my yogipartner and awesome paradise sharer, thank you for our awesome times together, and thank you for opening my eyes! I was wrong about American people (well maybe not most of them though, 'cause you're special), and now I have the

feeling I might understand them better (which is good as I will probably move to the US?!) so thanks! Ontzettend veel dank en liefde ook voor mijn **PSB Quiz** team met wie ik de laatste drie jaar mijn sterallures en quizkunsten heb mogen showen (**Rudy**, wees gerust ik vertrek enkel fysiek, mentaal blijf ik jou achtervolgen! Let jij maar goed op **Hendrik**, zodat ik jullie nog veel samen mag komen bezoeken! **Jenny,** spreken we af om samen **Stephen** eens te gaan bezoeken?!). Dank ook aan **Daniel Van Damme** voor al de hulp in de eerste fasen van het doctoraat, en voor het meerdere malen uitleggen van zovele plantenzaken waar onze groep minder ervaring mee had. Dank ook aan **Bert de Rybel:** voor de slaapzak, de droge humor en uiteraard de leuke en vooral goede experimenten die we samen hebben uitgevoerd (lees: die jij hebt uitgevoerd terwijl ik flauwe moppen tapte (of was het TAPte?)). Bedankt ook **Kirin**, voor al die keren dat je mij met de mais planten uit de nood hielp! En **Nubia** en **Liesbeth** voor de hulp met het herhalen van hun experimenten.

Ok, even ademhalen. Ik heb het gevoel dat ik nog niet eens halfweg zit, hahaha. Ik kan echt iedereen van het instituut hier beginnen opnoemen, zoals de mensen van de **IT groep**, die mij al zien komen van ver als ik weer eens met 'mijnen Apple' lastig kom doen. Ik ben een betere computernerd geworden dankzij jullie, hoera! Dank voor alle hulp! Of **Kristof Verleye**, die mij leerde solderen, mijn schoenen hielp plakken, .. Misschien moeten we de naam Bar Bricolage boven jouw deur hangen? Bedankt in elk geval voor de technische skills! En al de rest van de mensen die support geven, zoals **Carine** en het **maistransformatieteam, Miguel, Nico** en **Thomas** die voor de planten zorgen, dank jullie wel voor jullie tijd en hulp! Bedankt aan alle labovriendjes, zij die er nog zijn en zij die helaas al weg zijn, ik heb mij nog nooit zo goed thuisgevoeld op mijn werk als de laatste 5 jaar PSB.

A big thank you also for **Roman Pleskot,** for being so crazy to help me make my cover in such a short time, trying and trying again until it became so nice. You have a great talent there, nice to see other creative scientists in the world! Thanks a million for helping me make my thesis look like "me". Grote dank ook voor de rest van de kunstenaars van wie ik een kunstwerk mocht gebruiken voor mijn thesis! Bedankt **Simon** (eigenlijk was dit jouw idee, of heb je mij in elk geval toch kunnen overtuigen om buiten de standaardlijnen te kleuren en mijn network te gebruiken om er 'mijn' werk van te maken), merci **Miguel** voor de collage, en danku **Ciska** voor de mooie kei! Ik ben blij om zoveel kunstenaars rondom mij te hebben, want kunst en wetenschap zijn in mijn visie van de wereld heel nauw aan elkaar verbonden, en beiden zijn het hoogste menselijke goed.

En nu mijn bedankingen voor mensen die niet in onze labos rondlopen (pas op, het zijn er uiteraard nog veel meer, dus dit kan lang worden). Laat ik al beginnen met te zeggen dat ik jullie allemaal graag zie, en dat ik hoop dat ik er niemand ben vergeten tussen te zetten (maar indien dit wel het geval is, no worries, ik heb Alzheimer-op-jonge-leeftijd als excuus en een familiale verzekering, we regelen dat wel). Eerst en vooral mijn **ouders, broer en zus**: bedankt voor de steun dit jaar, dank dat ik thuis mocht komen schrijven want anders was het mij niet gelukt! Weet dat als ik binnenkort vertrek ik dit niet doe om bij jullie weg te zijn, maar omdat mijn kleine teen nu eenmaal kriebelt en weer naar het buitenland wil! Ik hoop jullie daar met veel liefde te mogen ontvangen! Sylke, dat geldt zeker voor jou en die lieve kleine neef van mij (en uiteraard zijn papa), want wat zou die zonder mij beginnen? Alleen op zijn hoofdstaan is nog geen goed idee, daar heeft ie mij voor nodig. Jonas, ik kom

zeker terug voor Boom, hint hint… Ook de rest van mijn familie wil ik graag bedanken, voor alle steun tijdens de afgelopen jaren.

Marc, mijn 'foutje van de tijd', mijn leermeester, medefilosoof, begripvolle luisteraar, en mijn (zonder twijfel) beste vriend. Danku vanuit het diepst van mijn hart om in mij te geloven, om mij te helpen bouwen aan een interessant leven, om zoveel deuren te open en mij zoveel dingen te leren kennen. Ik ben de gelukkigste mens op aarde met zulke wijze en lieve mensen rondom mij. Ergens wou ik dat ik voor altijd bij jou en **Nora** kon blijven, maar zoals dat moet met elk kuiken moet ook ik mijn vleugels uitslaan. Gelieve mij altijd te blijven overladen met boeken en geschreven wijsheid in eender welke taal!

Ook bedankt aan al mijn vrienden om mijn leven naast het werk zo leuk in te vullen. Bedankt voor de reizen, feestjes, BBQs, festivals, uitstappen, caféavonden, spelletjesavonden, diners, parksessies … Merci **Tineke,** dat ik bij jou al al die jaren altijd terecht kan! Je bent altijd welkom, waar ik ook zit! Veel success nog met je huisje en je carriere! Danku **Ciska,** voor de mooie jaren dat we samen woonden, bijna zusjes waren, Schevemieke en Ravemieke konden spelen, dat we samen met circus begonnen, en niet te vergeten, voor jouw ontzettend mooie kunst. Heel veel liefde!! Bedankt ook **Tobias** en **Gus**, mijn 'broertjes'. Ik moet zelfs niet zeggen waarom, you guys…! **Arne,** ik heb u ontzettend veel gemist, maar je bent loeihard voor je carriere aan het gaan en ik kijk vol verwondering toe! Goed bezig! Maar kom mij toch maar eens opzoeken in de toekomst, want van werken alleen word je niet gelukkig (en ik kan het weten na het schrijven van dit boek)! Dank ook aan alle mensen van de **Circusplaneet**, ik weet niet hoe ik mijn docoraat had overleefd zonder al die acrobatie, zonder hoepelen, zonder doek,.. Mijn innerlijke clown is tevreden en zal ik voor altijd blijven stimuleren! Al de rest van mijn vriendjes en mensen waarmee ik mijn leven en hart heb mogen delen: Ik ben geworden wie ik ben dankzij elke kleine ervaring met jullie! Voor zowel de leuke als de minder leuke, danku! Ik ga jullie missen als ik weg ben!! Vergeet mij niet, ok? Ik denk dat ik hard genoeg mijn best heb gedaan de laatste jaren om ervoor te zorgen dat dat nooit gebeurt! Dus, kom mij eens bezoeken en vooral: Keep on spreading the good Karma! Want Karma zaait ge altijd op voorhand!

Peace, Love, Science and Respect,
Uw wervelwind (Nienke)

Examination Board

Promoter

Prof. Dr. Geert De Jaeger	Centre for Plant Systems Biology, VIB Ghent University Belgium

Examination committee

Prof. Dr. Sofie Goormachtig (chair)	Centre for Plant Systems Biology, VIB Ghent University Belgium
Dr. Hilde Nelissen (secretary)	Centre for Plant Systems Biology, VIB Ghent University Belgium
Prof. Dr. Els Van Damme	Dept. Molecular Biotechnology, Ghent University Belgium
Dr. Jelle Van Leene	Centre for Plant Systems Biology, VIB Ghent University Belgium
Dr. Maarten Dedecker	AgroSafve, Ghent Belgium
Prof. Dr. Vicente Rubio	Centro Nacional de Biotecnología-CSIC, Spain

x

TABLE OF CONTENTS

ABBREVIATIONS XIII

SUMMARY XVII

SAMENVATTING XIX

SCOPE 1

CHAPTER 1: 5

FROM GENOTYPE TO PHENOTYPE: AN INTRODUCTION TO FUNCTIONAL GENOMICS

CHAPTER 2: 33

LEAF GROWTH REGULATION AND THE ROLE OF ANGUSTIFOLIA3 IN *ARABIDOPSIS* AND *ZEA MAYS*

CHAPTER 3 65

GS$^{YELLOW/GREEN}$: A MULTIFUNCTIONAL, "SWISS-KNIFE" TAP TAG FOR FUNCTIONAL PROTEIN ANALYSIS IN MONOCOTS AND DICOTS

CHAPTER 4 111

INSERTION OF THE TAP TAG IN THE GENOMIC CONTEXT OF A GENE THROUGH TAC RECOMBINEERING

CHAPTER 5 137

PULL-DOWN VERSUS TAP-MS: DO YOU NEED TWO TO TANGO?

CHAPTER 6 165

A PILOT STUDY TOWARDS THE IDENTIFICATION OF DIRECT TARGETS OF ZMAN3 IN DIVIDING AND EXPANDING TISSUES

DICUSSION AND PERSPECTIVES 195

CURRICULUM VITAE 205

SUPPLEMENTAL DATA 209

ABBREVIATIONS

AGO1	ARGONAUTE1
AIL	AINTEGUMENTA-like
ALC	ALCATRAZ
AmpR	Ampicillin resistance marker
AN3	ANGUSTIFOLIA 3
ANT	AINTEGUMENTA
AP	Affinity-purification
AP1	APETALA1
AP2	APETALA2
APC/C	Anaphase-Promoting Complex/Cyclosome
ARGOS	AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE
ARL	ARGOS-LIKE
ARP	ASYMMETRIC LEAVES 1/ROUGH SHEAT2/PHANTASTICA
ARP4	ACTIN RELATED PROTEIN4
AS1	ASYMMETRIC LEAVES 1
AS2	ASYMMETRIC LEAVES2
AUR1	AURORA1
AUX	Auxin
AUX1	AUXIN RESISTANT1
BAC	Bacterial Artificial Chromosomes
BAM3	BARELY ANY MERISTEM3
BAR	Basta resistance Marker
BB	BIG BROTHER
BIBAC	Binary Bacterial Artificial Chromosomes
BP	BREVIPEDICELLUS
BR	Brassinosteroid
BRG1	BRM-RELATED GENE1
BRI1	BRASSINOSTEROID INSENSITIVE1
BRM	BRAHMA
CaMV p35S	35S promoter of cauliflower mosaic virus
CCS52	CELL CYCLE SWITCH52
CDC27a	Cell division cycle protein 27 homolog a
CDK	Cyclin-dependent kinases
CDS	Coding sequence
CE-MS	Capillary electrophoresis–MS
ChIP-seq	Chromatin immunoprecipitation coupled to next generation sequencing
Chl	Chloroplast
CK	Cytokinin
CLIP	Cross-linking immunoprecipitation
CLV1-3	CLAVATA1-3
Co-IP	Co-immunoprecipitation
COL5	CONSTANS-LIKE5
CRF2	CYTOKININ RESPONSE FACTOR2
CRISPR/Cas9	Clustered, regularly interspaced, short palindromic repeat (CRISPR)/CRISPR associated protein9 (Cas9)
CYC	Cyclin

CYCB1;1	CYCLINB1;1
cYFP	Citrin Yellow Fluorescent Protein
CZ	Central zone
DAS	Days after sowing
DB	Differential Binding
DBD	DNA-binding domain
DCL4	DICER-LIKE4
DP	DIMERISATION PROTEIN
DREB2A	Dehydration-responsive element binding protein2A
DSBR	Double stranded DNA breaks
EMSA	Electrophoresis mobility shift essay
ER	Endoplasmic reticulum
ETT	ETTIN
EXP10	EXPANSIN10
FAE4	FASCIATED EAR4
FDR	False Discovery Rate
FE	Fold Enrichment
FI-ICR–MS	Fourier transform ion cyclotron resonance–MS
FIL	FILAMENTOUS FLOWER
FRT	Flippase recognition targets
FT	Unbound fraction
GA	Gibberellic acid
GA20OX1	Gibberellin 20-oxidase1
GC-MS	Gas chromatography-MS
GFP	Green fluorescent protein
GIF	GRF INTERACTING FACTOR
GO	Gene Ontology
GOF	Gain-of-function
GRF	GROWTH REGULATING FACTOR
HA tag	Hemagglutinin tag
HB-ZIPIII	Class III homeodomain-leucine zipper
HB33	HOMEOBOX33
HDT	Histone deacetylase
HEC1	HECATE1
HR	Homologous recombination
ICAT	Isotope-coded affinity tag
IGV	Integrative Genomics Browser
IPT7	Isopentenyl transferase7
KAN	KANADI
Kd	Dissociation constant
KLU	KLUH/CYP78A5
KN1	KNOTTED1
KNOX1	KNOTTED1-like homeobox protein
KRP	KIP-RELATED PROTEINS
LBL1	LEAFBLADELESS1
LC-MS	Liquid chromatography–MS
LFQ	Label-free quantification
LOF	Loss-of-function
MACS	Model-based analysis of ChIP-seq
mito	Mitochondria

MRM	Multiple reaction monitoring
MS	Mass spectrometry
MSCA1	MALE STERILE CONVERTED ANTHER1
NGS	Next-generation sequencing
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance spectroscopy
NSAF	Normalised Spectral Abundance Factor
ORF	Open reading frame
ORS1	ORGAN SIZE RELATED1
P0	Incipient leaf primordium
PAC	Phage 1 Artificial Chromosomes
phab	*phabulosa*
PIKK	phosphatidylinositol 3-kinase-related protein kinase
PIN1	PIN-FORMED1
PLA1	PLASTOCHRON1
PPD	PEAPOD
PPD2	PEAPOD2
PPI	Protein-protein interaction
Psr1;GARP	Golden2/*Arabidopsis* response-regulator/O starvation/acclimatisation response
pUBIL	Ubiquitin promoter of maize
PZ	Peripheral zone
qTAP-MS	Quantitative tandem affinity purification coupled to mass spectrometry
RBR	RETINOBLASTOMA-RELATED PROTEIN
RDR6	RNA-DEPENDENT RNA POLYMERASE
REV	REVOLUTA
RGD2	RAGGED SEEDLING2
rhino	Human Rhinovirus 3C protease cleavage site
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RNAi	RNA interference
RS2	ROUGH SHEAT2
RZ	Rib-zone
S6K	S6 kinase
SAGE	Serial analysis of gene expression
SAM	Shoot Apical Meristem
SAUR	Small Auxin-up RNA
SBP	Streptavidin-binding peptide
SCF	Skpl–cullin–F-box-protein E3 ubiquitin ligase
SEP3	SEPALLATA3
sgRNA	short guide RNA
SGS3	SUPPRESSOR-OF-GENE SILENCING3
SILAC	Stable isotope labelling with amino acids in cell culture
SLM	SCHLAFMUTZE
SLP1	SLEEPY1
SNP	Single nucleotide polymorphism
SPL	SQUAMOSA promoter binding protein-like
STM	SHOOTMERISTEMLESS
SWI/SNF	SWITCH2/SUCROSE NONFERMENTING2
SWP	STRUWWELPETER

SWP73	SWI/SNF ASSOCIATED PROTEINS73
SYT	Synovial sarcoma translocation protein
T-DNA	Transfer-DNA
ta-siRNA	Transacting small interfering RNA
TAC	Transformation Competent Artificial Chromosomes
TAD	Trans-activation domain
TAP	Tandem affinity purification
TChAP	Tandem Chromatin Affinity Purification
TCP	TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1 and 2
TD1	THICK TASSEL DWARF1
Tel2	Telomere maintenance2
TEV	Tobacco Etch Virus
TF	Transcription factor
TIC	Total Ion Current
TILLING	Targeting induced local lesions in genomes
TOR	TARGET OF RAPAMYCIN
TPX2	TARGETING PROTEIN FOR XENOPUS KINESIN-LIKE PROTEIN2
TPXL2	TPX2-like2
TLS	Translation Start Site
UVI4	ULTRAVIOLET-B-INSENSITIVE4
WOX	WUS-RELATED HOMEOBOX
WT	Wild type
WUS	WUSCHEL
Y1H	Yeast One-hybrid
Y2H	Yeas Two-hybrid
YAB	YABBY
YAC	Yeast Artificial Chromosome
ZAR1	Maize ARGOS1
ZHD5	ZINC FINGER HOMEODOMAIN5

SUMMARY

Proteins are one of the key components in the complex regulatory networks that underlie biological processes. Although some work independently, most proteins form interactions with other proteins or macromolecules, such as RNA and DNA, to perform their function. To aid functional annotation in plants in a system-wide manner, a wide variety of techniques have been constructed over the years (**Chapter 1**), which often imply a genetically encoded tag to avoid the need of protein specific antibodies. Two of the most powerful approaches that use this tagging principle are tandem affinity purification (TAP) followed by mass spectrometry to identify protein-protein interactions, and chromatin immunoprecipitation coupled to next generation sequencing to assess protein-DNA interactions. A third application of protein tagging is found in the visualisation of localisation patterns of proteins by means of fluorescent protein tags. However, no tags have been constructed in plants that combine all three approaches. This makes studying complex biological processes, such as leaf growth and development, in which a myriad of proteins play a regulatory role (**Chapter 2**), rather difficult.

In this research, we constructed and evaluated a "Swiss-knife" TAP tag, named GS^{yellow}, that allows the identification of interaction partners and localisation patterns of a protein under study (**Chapter 3**) in both *Zea mays* and *Arabidopsis thaliana*. By combining data gathered from TAP, ChIP and localisation experiments, we confirmed previously reported mechanisms of a set of bait proteins in their developmental contexts. Hence, the obtained data indicate that the GS^{yellow} tag performs equally well as other tags in the analysis of protein functions in plants. Furthermore, to preserve the endogenous regulatory sequences surrounding the gene of interest in *Arabidopsis*, an alternative workflow to produce genetically TAP-tagged fusion proteins based on homologous recombination in combination with transformation-competent artificial chromosomes (TAC), also called recombineering, was successfully implemented (**Chapter 4**). Moreover, to provide a deeper insight into the differences obtained with TAP and pull-down experiments, protein complexes purified from *Arabidopsis* and *Zea mays* with both techniques were compared (**Chapter 5**). This revealed that both experimental set-ups can detect different types of interactions, and that having both possibilities at hand provides a great flexibility for protein-protein interaction studies.

Finally, the applicability of the GS^{yellow} TAP tag in the functional analysis of a protein complex in a developmental context was demonstrated in *Zea mays* leaves via the transcriptional co-activator ANGUSTIFOLIA3 (ZmAN3) that plays a known regulatory role in leaf growth (**Chapter 6**). Here, previous TAP data of ZmAN3 that indicated preferential binding with transcription factors of the GROWTH REGULATING FACTOR (GRF) family in the dividing tissues of the growth zone was confirmed. Additionally, target gene identification through ChIP-seq on the full

growth zone as well as on the division and expansion zone separately showed that ZmAN3 regulates the expression of genes involved in the maintenance of the proliferative state. Our data also indicated that the regulatory feedback loop that regulates GRF expression through binding of ZmAN3-GRF to its own promoter, is conserved between dicots and monocots.

In conclusion, the multifunctional GS^{yellow} TAP tag that was evaluated here can be used to analyse the function of protein complexes in a developmental context in both monocots and dicots. By analysing the function of the growth regulatory ZmAN3 protein in leaf development, it was demonstrated how the GS^{yellow} tag can be applied to gain insights into complex biological processes, which can aid to engineer the underlying pathways more precisely in the future.

SAMENVATTING

Eiwitten zijn een van de belangrijkste spelers in de complexe regulatorische netwerken die aan de basis liggen van vele biologische processen. Hoewel sommige onafhankelijk kunnen werken, hebben de meeste eiwitten interacties met andere eiwitten of macromoleculen, zoals RNA en DNA, nodig om hun functie uit te oefenen. Om het functioneel analyseren van planteneiwitten te vergemakkelijk werden er in het verleden meerdere technieken ontworpen (**Hoofdstuk 1**), die vaak een genetisch geëncodeerde tag implementeren om het gebruik van eiwit-specifieke antilichamen te vermijden. Twee zeer krachtige toepassingen die dergelijk taggingsysteem toepassen zijn tandemaffiniteitszuivering (TAP) gevolgd door massaspectrometrie om eiwit-eiwit interacties te identificeren, en chromatine immunoprecipitatie (ChIP) gekoppeld aan de nieuwste generatie sequeneringsmethoden om eiwit-DNA interacties te bepalen Een derde toepassing voor eiwittagging is gelinkt met het visualiseren van de lokalisatiepatronen van een eiwit door middel van fluorescente eiwittags. Er bestaan echter geen tags in planten die deze drie technieken combineren. Dit maakt het bestuderen van complexe biologische processen zoals bladgroei en onwikkeling, waarbij een grote waaier aan regulatorische eiwitten een rol spelen (**Hoofdstuk 2**) zeer moeilijk.

Dit onderzoek focuste zich op het ontwerpen en evalueren van een multifunctionele TAP tag, genaamd GS^{yellow}, die kan gebruikt worden om interactiepartners en lokalisatiepatronen te identificeren van een bepaald eiwit in zowel *Arabidopsis thaliana* en *Zea mays* (**Hoofdstuk 3**). Door data bekomen via TAP, ChIP en lokalisatieëxperimenten te combineren konden we eerder gerapporteerde mechanismen van een selecte set eiwitten bevestigen in hun ontwikkelingscontext. Onze data tonen dus aan dat de GS^{yellow} tag even goed presteert als andere tags die gebruikt worden in functionele eiwitanalyse in planten. Daarnaast hebben we ook een alternatieve werkwijze opgesteld in *Arabidopsis* om genetische TAP-getagde fusieëiwitten te maken waarbij de endogene regulatorische sequenties die rond het gen van interesse liggen worden behouden (**Hoofdstuk 4**). Hiervoor werd gebruik gemaakt van homologe recombinatie in combinatie met transformatie competente artificiële chromosomen (TAC), een process dat beter bekend staat als recombinering.
Om een dieper inzicht te krijgen in de verschillen die bekomen worden met TAP en 'pull-down' experimenten werden er bovendien een aantal complexzuiveringen uitgevoerd in zowel *Arabidopsis* als *Zea mays* met beide technieken (**Hoofdstuk 5**). Door de resultaten te vergelijken werd het duidelijk dat beide experimentele methoden verschillende interactietypes kunnen oppikken, en dat de mogelijkheid om te kiezen tussen beide opties met de GS^{yellow} tag een grote flexibiliteit levert in eiwit-eiwit interactiestudies.

Tot slot werd de bruikbaarheid van de GS^{yellow} TAP tag in de functionele analyse van eiwitcomplexen in een ontwikkelingscontext aangetoond in *Zea mays* (**Hoofdstuk**

6). Hierbij werd gebruik gemaakt van de transcriptionele co-activator ANGUSTIFOLIA3 (ZmAN3), welke een gekende regulatorische rol speelt in bladgroei. De vroeger bekomen TAP data van ZmAN3, waarbij aangetoond werd dat het eiwit preferentieel bindt met transcriptiefactoren van de GROWTH REGULATING FACTOR (GRF) familie in de delende weefsels van de groeizone, werd hier bevestigd. Bovendien toonde de identificatie van doelwitgenen via ChIP-seq, op zowel de volledige groeizone als op de delende en expanderende weefsels apart, dat ZmAN3 de expressie regelt van genen die betrokken zijn in het behoud van de proliferatieve eigenschappen van de groeizone. Onze gegevens duiden ook aan dat de regulatorische feedback loop, waarbij GRF expressie geregeld wordt door binding van ZmAN3-GRF aan zijn eigen promotor, geconserveerd is tussen monocotyle en dicotyle planten.

De multifunctionele GS^{yellow} TAP tag die hier geëvalueerd werd kan dus worden gebruikt om de functie van eiwitcomplexen te bestuderen in een ontwikkelingscontext in zowel monocotyle als dicotyle planten. Door de functie van het groeiregulatorische ZmAN3 eiwit in bladgroei te analyseren werd aangetoond hoe de GS^{yellow} tag kan worden toegepast om een inzicht te vergaren in complexe biologische processen. Dit kan ons in de toekomst helpen voor de preciezere engineering van de onderliggende pathways.

Scope

The increase in publicly available whole-genome sequences since the introduction of Next Generation Sequencing (NGS) has created a platform for system-wide biological studies. In an attempt to annotate all these new genes and their gene products, the field of functional genomics combines genome- and system-wide analysis of DNA sequences, gene expression and protein functions, and tries to gain an insight into complex biological processes, such as development and growth. As protein functions are predominantly determined by their interaction partners, identification of these interactors (*i.e.*, proteins, DNA, RNA or metabolites) has become of major interest.

One of the most powerful biochemical approaches to identify protein-protein interactions in a proteome-wide manner is Tandem Affinity Purification (TAP) coupled to mass spectrometry (MS). Based on the production of a recombinant protein tagged with a double-affinity tag that allows two subsequent rounds of affinity purification under physiological conditions, the technique provides very pure protein complexes for downstream analysis. As such, co-purifying interaction partners can easily be identified through MS with high throughput. Besides TAP, other protein analysis techniques including Bimolecular Fluorescence Complementation (BiFC) and pull-down, also often use tagging to produce fusion proteins through recombinant expression. This has led to the development of a wide variety of epitope and protein tags in the past. Moreover, protein tags, such as the Green Fluorescent Protein (GFP), are frequently implemented to analyse the tissue specificity and subcellular localisation of gene products. A third application of affinity tags can be found in the determination of target genes of transcription factor through Chromatin Immunoprecipitation (ChIP). Protein tagging can as such be used not only to isolate protein complexes, but also to analyse the function of these proteins. Unfortunately, no tags have been constructed in plants that combine the different applications of protein tagging into one tag, which often leads to an increase in workload of producing transgenic lines for each gene construct and makes direct comparison of the gathered results on all these lines difficult.

Therefore, this doctoral research was set up to provide a technology that allows the isolation and characterisation of protein complexes in their developmental context in plants. By developing a multifunctional TAP tag that allows both purification of protein complexes as well as the identification of different types of binding partners and localisation patterns, a more in-depth understanding of the function of the complex under study can be obtained. As such, studying protein complexes in their developmental context in plants will become more straightforward.

The first aim of this project is the construction and evaluation of a "Swiss knife" TAP tag that combines all these approaches. By selecting a handful of model proteins, the new TAP tag will be evaluated in *Arabidopsis thaliana* plants, and protocols will be developed for each of the different applications. Next, the technology will be transferred to corn, which is a useful model plant to analyse the function of growth regulatory complexes in a developmental context. To this end,

a single gene is selected that provides an increased leaf size upon overexpression in *Arabidopsis* and is known to be incorporated into a protein complex.

The second goal of the project is to evaluate a different method to construct recombinant proteins in *Arabidopsis* with the emphasis on the conservation of the endogenous context of the gene of interest. In model species, such as yeast and *Escherichia coli,* genetic tagging is often achieved through knock-in of the tag sequence via homologous recombination (HR). However, the efficiency of HR in plants is very low. Therefore, a strategy will be tested that allows HR of the new TAP tag into artificial chromosomes to preserve the 5' and 3' endogenous context of the gene of interest. As the endogenous landscape forms an important part in gene regulation, providing such a HR-based tool to maintain this biological information during functional protein analysis is both useful and necessary. For this purpose, the same set of bait proteins that were selected to evaluate the performance of the new TAP tag will be used to allow comparison of the obtained results.

A third aim of the project is to provide a deeper insight into the differences gathered with pull-down and TAP experiments, which are two of the most commonly applied techniques to study protein complexes in plants. Using the data obtained in the first part of this project as a reference together with some additional bait proteins, a comparison will be made in both *Arabidopsis* cell cultures and seedlings, as well as in the developmental stages of the growth zone of the maize leaf. Here, the new tag will be an advantage, as both techniques can be carried out on the same biological material, allowing direct comparison of the gathered results.

In the final part of this project, we will demonstrate the usability of the new TAP tag in the functional analysis of growth regulatory protein complexes in a developing leaf of *Zea mays*. Due to the small size of the *Arabidopsis* leaf during the earliest developmental phases, studying leaf growth is rather hard in the latter. The larger leaf organs of the maize plant, on the other hand, are a better tool for this, and maize plants are also of high value for agriculture. By using the new TAP tag developed in the first part of the project, the protein complex will be characterised in different developmental tissues along the growth zone of the leaf. The data collected from localisation assays, complex purifications and ChIPs on these different tissues in both wild-type and transgenic plants will provide a deeper insight on the function of the protein complex during leaf growth. The idea is to obtain as much information as possible to engineer leaf growth more precisely in the future.

FROM GENOTYPE TO PHENOTYPE: AN INTRODUCTION TO FUNCTIONAL GENOMICS

Nienke Besbrugge[1,2] and Geert De Jaeger[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2]Centre for Plant Systems Biology, VIB, 9052 Ghent, Belgium

Author contributions:

N.B. wrote the chapter; G.D.J. supervised and complemented the writing

ABSTRACT

The introduction of next generation sequencing techniques has created a vast increase in the number of publicly available whole-genome sequences. In an attempt to annotate all these new genes and gene-products, the functional genomics field combines data gathered from genome- and system-wide analysis techniques to provide a deeper insight into complex biological processes. Here, an overview is given of the most commonly applied tools in genomic, transcriptomic, proteomic, interactomic and metabolomic experiments in plants. Starting with the identification of gene sequences, all aspects of unravelling the function of the gene product by integrating interaction and expression data are covered.

INTRODUCTION

The discovery of DNA in the second half of the 20th century as the genetic code that drives the diversity of life has transformed the world of life sciences. Since then, a remarkable series of scientific and technological advances have created a more in-depth view of living systems and provided a wide range of applications that are beneficial for society. Starting with the sequencing of the human genome, the field of genomics has grown tremendously thanks to the emergence of next-generation sequencing (NGS) tools (Bunnik and Le Roch, 2013). For the plant field, the *Arabidopsis thaliana* genome was the first to be completed (in 2000). At present, more than 180 plant genomes, including rice (Sasaki and Burr, 2000), maize (Schnable *et al.*, 2009), sorghum (Paterson *et al.*, 2009), poplar (Tuskan *et al.*, 2006), grapevine (Jaillon *et al.*, 2007), papaya (Chan, 2009), soybean (Schmutz *et al.*, 2010) and watermelon (Guo *et al.*, 2012) have been sequenced, and this number keeps increasing. Whereas sequencing costs and processing time are constantly decreasing, the accuracy of sequence determination has now even reached the point where single nucleotide differences between individual genomes can efficiently be detected. However, assembling the genome of a species is insufficient to understand how genes work and interact with each other in the complex networks that underlie biological processes, such as development and response to a changing environment (Rahmat *et al.*, 2016). Therefore, the functional genomics field tries to determine the function of genes and their interactions in biological pathways. By combining information gathered with transcriptomic, proteomic, interactomic and metabolomic experiments, it aims to provide a deeper insight at the systemic level rather than at the single-gene level (Figure 1) (Ben-Amar *et al.*, 2016). As such, typical functional genomics tools are large scale, providing a genome- or system-wide coverage with high throughput. Although a wide variety of techniques have been developed in the past, only the most common will be discussed here.

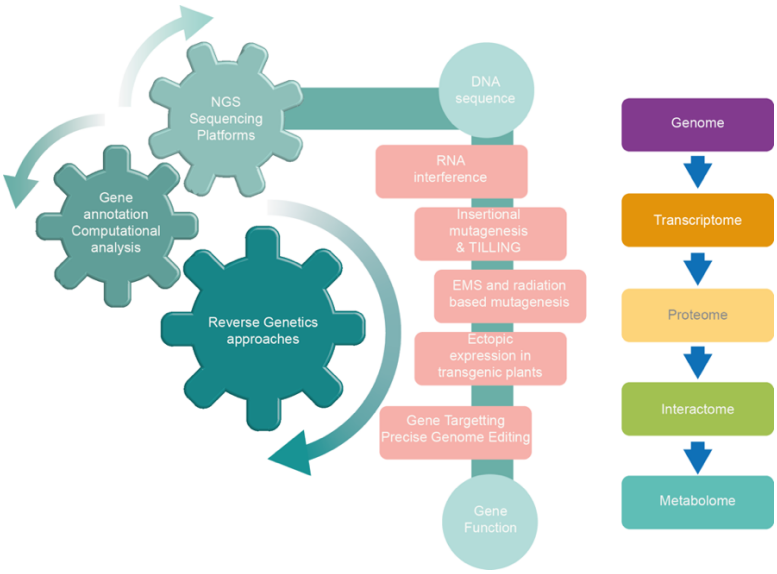


Figure 1. Overview of reverse genetics approaches and the different levels of analysis in functional genomics that are covered in this chapter. Figure adapted from (Ben-Amar *et al.*, 2016)

GENOMICS

The first generation of sequencing methods was based on the Sanger dideoxynucleotide sequencing method, which requires processing of each sequence read separately (Heather and Chain, 2016). Despite several attempts to increase automation and throughput of the technique over the years, it is still a laborious and expensive process. Moreover, since whole-genome sequencing requires the production of bacterial clones, cloning of AT-rich fragments or genes toxic to the bacteria can create biases (Bunnik and Le Roch, 2013). As such, the field of genomics has undergone a transformation with the commercialisation of NGS methods. With up to 6 million reads or 600 gigabase per instrumental run being processed in parallel, the amounts of data originating from small DNA samples are enormous (Schwamb and Pallasch, 2014). Mutations, such as single-nucleotide polymorphisms (SNPs), deletions or duplications can now easily be identified with whole-genome sequencing. Moreover, due to an increasing throughput and a reduction in costs, NGS has been coupled to a wide range of novel techniques, including analysis of DNA methylation, histone modification, transcription factor (TF) binding, nucleosome positioning and other epigenetic mechanisms along the genome via Chromatin Immunoprecipitation coupled to NGS (ChIP-seq)(Qureshi and Mehler, 2011). Another application of NGS can be found in the determination of gene expression profiles through the sequencing of transcripts (see Transcriptomics). Both methods require mapping of short sequence reads to a reference genome, and analysis of read distributions. Binding sites are mapped from the flanking regions and can now even be determined with single-nucleotide

resolution. With the introduction of third-generation sequencing methods, which can process DNA molecules singularly without the need for amplification of template DNA, sequencing has never been cheaper whereas offering high throughput (Heather and Chain, 2016).

Once the genome is sequenced and assembled, gene functions are predicted *in silico* using bioinformatics tools (Bolger *et al.*, 2017). To this end, proteomes and genomes are compared via homology with databases that contain nucleotide and amino acid information of different species (Abdeeva *et al.*, 2012). To identify candidate genes, genomes can be scanned for long open reading frames (ORFs), sequences linked to initiation of transcription, polyadenylation sites or other regions that are linked to protein-coding genes. To validate the putative gene, downstream analysis is always needed. Several strategies can be followed to confirm such candidate genes, including searching for resemblance of the gene with already characterised cDNA sequences within the same organism, finding homology of the predicted protein sequence to known proteins, *etc* (Abdeeva *et al.*, 2012). To facilitate such comparisons, full-length (fl)-cDNA (also containing the 5' and 3' UTR) libraries of a wide range of plant species, such as rice (Kikuchi, 2003), maize (Sekhon *et al.*, 2013), cassava (Sakurai *et al.*, 2007) and sorghum (Abdel-Ghany *et al.*, 2016) have been made publicly available (Rahmat *et al.*, 2016). Databases containing millions of expressed sequence tags (ESTs, short sub-sequences of a cDNA sequence that can be used for the identification of gene transcripts) from several crop species as well as from transcriptomes obtained from different tissues, developmental stages or plants grown under different types of stresses, have also been constructed (Rahmat *et al.*, 2016). Besides their implementation in the identification of gene sequences, fl-cDNA collections also provide an important tool in the analysis of the gene's function (see below).

FORWARD GENETICS

As soon as gene sequences have been identified, their biological function can be studied. A general approach to do this is through mutagenesis of the candidate gene. Here, two strategies can be followed: forward genetics or reverse genetics. Forward genetics starts with the selection of a biological process of interest. Next, a wide variety of mutant plants are generated, and plants providing a desired phenotype are used to map and clone the gene that causes the phenotype. It has the advantage that new genes can be discovered without the necessity of knowing the gene's function *a priori* (Alonso and Ecker, 2006). In *Arabidopsis*, forward genetics has been extensively used to unravel the genes involved in most biochemical processes and signalling pathways (Rahmat *et al.*, 2016). For example, ethylene-related genes that function in the production, signalling, perception or responses to ethylene were identified using forward genetics. Hence, mutations were mapped that caused a specific morphological change upon ethylene treatment in dark-grown seedlings (Gichner, 1995; Rahmat *et al.*, 2016). Common mutagenesis techniques that are used in forward genetics are exposure to chemicals (EMS) or radiation with fast neutrons, and T-DNA or transposon insertions. Based on these methods, several collections of mutants have been made

available for species, such as *Arabidopsis* (The *Arabidopsis* Information Resource; http://www.arabidopsis.org) and rice (International Rice Functional Consortium; http://irfgc.irri.org) (M Perez-de-Castro *et al.*, 2012), which provide different types of ready-to-screen mutagenized seeds for functional analysis. With these collections at hand, system-wide forward genetics that analyse the phenotype of all gene-indexed mutants at the same time has become a lot easier.

REVERSE GENETICS

The availability of so many sequenced genomes, on the other hand, has created a platform for reverse genetics, also known as the 'genotype-to-phenotype' approach. Here, known gene sequences are mutated through molecular techniques to see whether they produce a phenotype (Shalem *et al.*, 2015). One of the advantages of the reverse strategy is the ability to analyse genes belonging to gene families. As these often have redundant functions and fail to provide a phenotype upon mutation, they cannot be identified using forward genetics. In general, the gene of interest is perturbed to produce either a loss-of-function (LOF) or a gain-of-function (GOF) mutant. Therefore, molecular techniques, such as RNAi-mediated post-transcriptional gene silencing (for instance, The AGRIKOLA project (http://www.agrikola.org) in *Arabidopsis*) (Senthil-Kumar *et al.*, 2010), T-DNA insertion mutagenesis, or TILLING (Targeting induced local lesions in genomes), are being used to decrease gene expression (*i.e.,* LOF). TILLING induces point mutations in the genomes of, for example, a genetic stock, improved germplasm, *etc.*, via chemical or radioactive treatment, and detects the mutations through PCR (McCallum *et al.*, 2000) or NGS (Pan *et al.*, 2015). Compared to RNAi, T-DNA or transposon insertion, TILLING does not require the usage of genetic transformation techniques. This makes it very effective (Ronald, 2014), especially for plants for which transformation methods are still inefficient or unavailable (Alonso, 2003; Nakamura *et al.*, 2007; Rahmat *et al.*, 2016). A variant of TILLING, EcoTILLING, is also often used in multiple crop species to identify natural polymorphisms of alleles of genes in a species, as has been shown for the analysis of a number of TFs in rice that are involved in drought stress tolerance (Yu *et al.*, 2012). With the availability of large collections of mutants for which the mutations are well documented, reverse genetics approaches can easily be applied to unravel gene functions (Alonso and Ecker, 2006).

Increasing gene expression, on the other hand, can be achieved through activation tagging by T-DNA insertion or heterologous expression of the gene through a strong promoter (*i.e.* GOF)(Weigel *et al.*, 2000; Nakazawa *et al.*, 2003; Ichikawa *et al.*, 2006). Because only one gene is affected without perturbing the expression of other family members, functionally redundant genes can easily be studied (Ito, 2000; Nakazawa *et al.*, 2008), which is not the case with LOF mutants. Common strategies for overexpression in plants are the usage of the *cauliflower mosaic virus (*CaMV) 35S promoter in *Arabidopsis* (Abdeeva *et al.*, 2012) or the UBIL promoter (Karimi *et al.*, 2007) in *Zea mays* in combination with a gene's full length cDNA or ORF sequence (Abdeeva *et al.*, 2012).

TARGETED GENOME EDITING

Recent advances in the field of genomics have made it possible to precisely edit the genome by targeting insertions, deletions or changes at a genomic sites of interest via designed nucleases (Sander and Joung, 2014). Probably the most popular of these techniques is the recently developed clustered, regularly interspaced, short palindromic repeat (CRISPR)/CRISPR associated proteing (Cas9) system, which introduces double stranded breaks at the targeted site (Rinaldo and Ayliffe 2015). Thus, a short single guide RNA (sgRNA), complementary to the target DNA sequence, is introduced together with a Cas9 protein. In plants, the CRISPR/Cas9 system has become immensely popular, which is reflected in its implementation in several crop species, such as wheat, soybean and rice to mutate specific genes (Jacobs *et al.*, 2015; Shen *et al.*, 2017).

In conclusion, genome editing and mutagenesis are useful tools to get an idea of the function of a gene. However, to obtain a more in-depth view on the specific regulation of the gene, more information coming from transcriptomics or interactomics studies is needed.

TRANSCRIPTOMICS

The field of transcriptomics tries to capture the transcriptional profile of a genome under specific conditions or in a specific cell type by measuring all produced RNA transcripts. To this end, the expressed part of a gene, *i.e.*, the mRNA, is identified using high-throughput methods, such as microarray analysis or RNA-seq (Wang *et al.*, 2009). The technique is based on the fact that environmental or developmental cues induce transcriptional responses that are specific for certain types of cells (Gifford *et al.*, 2008; Dinneny *et al.*, 2008). Once the transcript profile of a given cell type or condition is known, it can help to unravel the function of the genes and the regulatory pathways that are involved in the response. Moreover, by comparing transcriptomes from distinct cell types or derived through different treatments, differentially expressed genes can be identified.

In general, gene expression profiling is often combined with LOF or GOF mutants, as this allows the exploration of the function and characteristics of a selected gene (Nakamura *et al.*, 2007). One of the first transcriptomic profiling methods is serial analysis of gene expression (SAGE) (Aldaz, 2003). By sequencing small tags that correspond to the 3' end of all mRNAs in a sample, it results in a so called snapshot of the RNA population of that sample. Although the technique allows the easy quantification of gene expression by counting all the tags for a given gene, its labour-intensive and relatively low throughput characteristics make it less attractive compared to NGS-based methods (van Ruissen *et al.*, 2005). Another strategy to analyse the transcriptome of a sample is based on the usage of DNA microarrays, which measure the hybridisation of labelled single-stranded cDNAs or cRNAs from a sample to spotted DNA probes on the array (Abdeeva *et al.*, 2012; Bunnik and Le Roch, 2013; Rahmat *et al.*, 2016). Here, the amount of hybridisation corresponds directly to the transcript level of the gene in the sample. The

implementation of microarrays for gene expression profiling is still very popular these days, but is limited to the detection of transcripts for which a probe is present on the array (Heller, 2002). RNA-seq, on the other hand, can detect any type of RNA (including long noncoding RNAs, nascent transcripts, splice variants etc.) in a sample, but also requires a reference genome or transcriptome to which the sequence reads are mapped (Wang *et al.*, 2009; M Perez-de-Castro *et al.*, 2012). RNA-seq is at present the most high-throughput transcriptomics tool that provides a genome-wide coverage together with a large sequencing depth. Besides its usability in gene expression profiling, RNA-seq also allows identification of novel transcripts or splice variants. Another application can be found in the investigation of the boundaries of a transcript and intron/exon junctions (Bunnik and Le Roch, 2013).

Although gene expression profiling is a useful technique to obtain knowledge on gene function, some considerations have to be taken into account. First of all, changes in transcript levels do not always result in changes in protein abundance (Mittler, 1998). Second, it is impossible to detect any post-translational modifications of proteins, which often influence a protein's activity. As such, the best approach to study gene expression is by combining transcriptomic techniques with proteomic analyses (Walther and Mann, 2010).

PROTEOMICS

Studying the roles of proteins, which are one of the functional units of cells, is crucial to understand complex biological processes. As the proteome of a given cell at a specific time point is more variable than the transcriptome of an organism, identification of the protein content and the correlating protein abundance is more challenging than transcriptomics approaches. In the past, protein levels were estimated from RNA analysis. However, transcript levels were found sometimes not to correlate to protein abundance because protein production is influenced by mRNA stability and the efficiency of translation. Moreover, several splice variants can be produced from a single transcript, and post-translational modifications can provide a distinct function to proteins. Given that some proteins can only function in the presence of interacting partners and that degradation rates of proteins can vary, proteomic methods provide a different type of information than genomic or transcriptomic analyses (Walther and Mann, 2010).

One of the first ways to assess a protein's function is by analysing its localisation pattern, because the chemical environment in organelles can alter the activity of a protein (Itzhak *et al.*, 2016). As such, combining compartmentalisation and protein movement data with protein abundance profiles can help to understand, for example, responses to biological cues. In general, two different strategies can be followed to obtain spatial information: either genetic tagging of proteins with a fluorescent marker and microscopy-based analysis of localisation, or mass-spectrometry (MS)-based measurement of cellular compositions in different conditions or in different developmental tissues (Aebersold and Mann, 2016). In plants, one of the most commonly applied fluorescent markers for microscopy-

based analysis is Green Fluorescent Protein (GFP)(Figure 2) (Collings, 2013; Dedecker *et al.*, 2015; Lim *et al.*, 2016). By combining data from large-scale subcellular proteomics, fluorescent protein visualisation and protein–protein interaction (PPI) studies with subcellular targeting calls from 22 prediction programs, a subcellular location database for Arabidopsis proteins has been made available at http://suba.live/. GFP is also a common tag in other model organisms, such as yeast, where it has been extensively used in combination with pull-down quantitative MS (also known as affinity enrichment (AE)-MS) to identify all interactors of a protein (see further) (Keilhauer *et al.*, 2014).

Despite the ease of studying the spatial dynamics of individual proteins via fluorescent markers, the technique has its limitations. First of all, the presence of the tag can interfere with the endogenous function of the protein by altering its ability to move or interact with other molecules. Second, expression of the transgene has to be regulated by the endogenous regulatory sequences to obtain a biological idea of localisation. This is rather difficult in plants, as homologous recombination events are rare and promoter and enhancer sequences that regulate transcription are still poorly understood (Zhou *et al.*, 2011; Dedecker *et al.*, 2015). Additionally, the throughput of tagging-based localisation analysis depends a lot on the transformation efficiency of the species of interest (Walther and Mann, 2010).

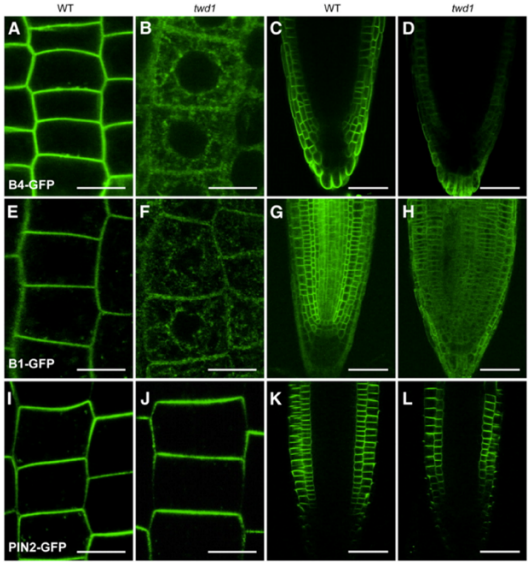


Figure 2. Analysis of tissue specificity and subcellular localisation via the Green Fluorescent Protein marker. Shown here for PIN2-GFP in cortical cells of *Arabidopsis* roots. Figure taken from (Wu *et al.*, 2010)

MS-based methods, on the other hand, are gaining popularity for large-scale cellular composition studies, as they allow both the identification and quantification of all proteins in a sample in a single experiment (Walther and Mann, 2010). Based on the measurement of peptide masses, where peptides are obtained through

enzymatic digestion of the full proteins, the technique also allows the identification of post-translational modifications. However, as measured peptides are compared to a database for identification of the protein from which the peptide originated, the technique can only identify predicted or already documented proteins. With constantly advancing technologies in the field including higher sensitivity of the MS as well single-shot LC-MS, MS-based proteomics has now completely replaced previous tools, such as two-dimensional gel-electrophoresis for protein content analysis (Meyer and Selbach, 2015; Itzhak *et al.*, 2016). With the implementation of single-shot LC-MS methods, extensive frationation of the sample can be avoided, which leads to a reduced handling time as well as higher accuracy and sensitivity. As such, it is nowadays feasible to assess entire model proteomes in a single LC-MS run (Nagaraj *et al.*, 2012).

QUANTITATIVE PROTEOMICS

Sometimes it is more interesting to assess changes in proteome composition under different conditions or in different cell types rather than knowing whether a protein is present or not. Therefore, quantitative proteomics, generally refered to as multiple or selective reaction monitoring (MRM or SRM) proteomics, is now becoming a standard in the proteomics field (Itzhak *et al.*, 2016). The MRM characteristics of the triple quadrupole MS instrument allows to select a uniquely daughter ion that is obained after fragmentation of a parent ion, and allows as such to monitor specific peptides. To measure the changes in peptide abundances, two strategies can be followed: isotope-labelling and label-free quantification (LFQ) (Meyer and Selbach, 2015).

A first option for isotope-labelling is SILAC (stable isotope labelling with amino acids in cell culture) (Figure 3) (Amanchy *et al.*, 2005), which uses growth media supplemented with labelled amino acids. As these amino acids are metabolically built into the protein, 'light' versus 'heavy' versions of the proteins are produced. Although the technique was originally established for cell cultures, it has also been used with living organisms, such as mice and *Caenorhabditis elegans* (Krüger *et al.*, 2008; Geiger *et al.*, 2010). Originally, SILAC-based metabolic labeling methods were unsuitable for plants due to the plant's autotrophic nature (Dedecker *et al.*, 2015). Nowadays, however, the protocol has been optimised also for *Arabidopsis,* and has helped to identify differentially expressed proteins upon salt stress in seedlings (Lewandowska *et al.*, 2013). Another labeling approach that circumvents the need for metabolic based labeling is isobaric tagging. Compared to metabolic labeling, isobaric tagging quantitates peptides in the MS2 spectra, as the labeled peptides will fragment into ions that directly represent the isotpopically labeled reporter regions. Hereto, collision induced dissociation (CID) is applied, which results in spectra that contain both quantiative and identification data for each peptide in the experiment in a single run. With the introduction of higher-energy collisional dissociation (HCD) in Orbitrap instruments, the accuracy of the techique to quantify peptides is constantly increasing. Although this higher accuracy comes at the price of longer scanning times and a lower number of detected peptides, the constantly improving reagent costs and reproducibility of the protocol are creating

an uprise in popularity of the technology (Köcher *et al.*, 2009). Two commonly applied techniques that use this approach are isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags. iTRAQ can be used with any type of sample, and is a popular technology in the field of clinical sample analysis (Evans *et al.*, 2012; Huang *et al.*, 2014; Zhang *et al.*, 2015). With the availability of plant specific protocols (Martínez-Esteso *et al.*, 2014; Handakumbura *et al.*, 2016), the methodology is now also gaining more attention in the plant field. For example, performing iTRAQ on soybean leaves and roots under salt treatments revealed that respectively 278 and 440 proteins were significantly altered. These proteins originated from 13 biological processes, showing that salt stress may influence these processes directly, and allows to narrow down downstream analysis in salt stress resistant plants (Ji *et al.*, 2016). Another study on red algea (*Pyropia haitanensis*) focused on the iTRAQ-based identification of proteins altered by different durations of high temperatures in a heat stress tolerant strain to gain insight into the coping mechanisms of the plant during heat stress (Shi *et al.*, 2017).

To circumvent the limitations of labelling, label-free quantification methods have been developed that are becoming more and more popular in proteomic studies (Dedecker *et al.*, 2015). These methods use multiple replicates to calculate differences in peptide peak intensities between distinct peptide mixtures via complex statistical or probability algorithms. Because differences between samples have to be large enough to be picked up, the accuracy of LFQ methods is generally lower than that of labelling-based methods. There are two major strategies for label-free quantification: spectral counting or quantitation based on extracted ion chromatograms (XIC). The first approach calculates differences via a normalised spectral abundance factor (NSAF), which counts the total number of identified spectra per protein and normalises this against the length of the protein and the total number of identified spectra in the sample (Law *et al.*, 2010; Dedecker *et al.*, 2015). The second approach counts peptide ions, and relies on the assumption that higher peptide concentrations result in a greater area of the chromatographic peak identified in MS spectra. As such, the peak area corresponds to the total ion current (TIC, *i.e.*, the sum of all ion counts) divided by the peptide elution time of all peptides identified in the sample. When peak areas of all peptides of a given protein are known for each sample, they can be compared to obtain a relative protein concentration. Thanks to the constantly improving computational methods, cleaning up the gathered data is becoming easier, which translates itself into an increasing popularity of LFQ methods. One of the best known algorithms that uses TIC is MaxQuant (see below)(Cox and Mann, 2008).

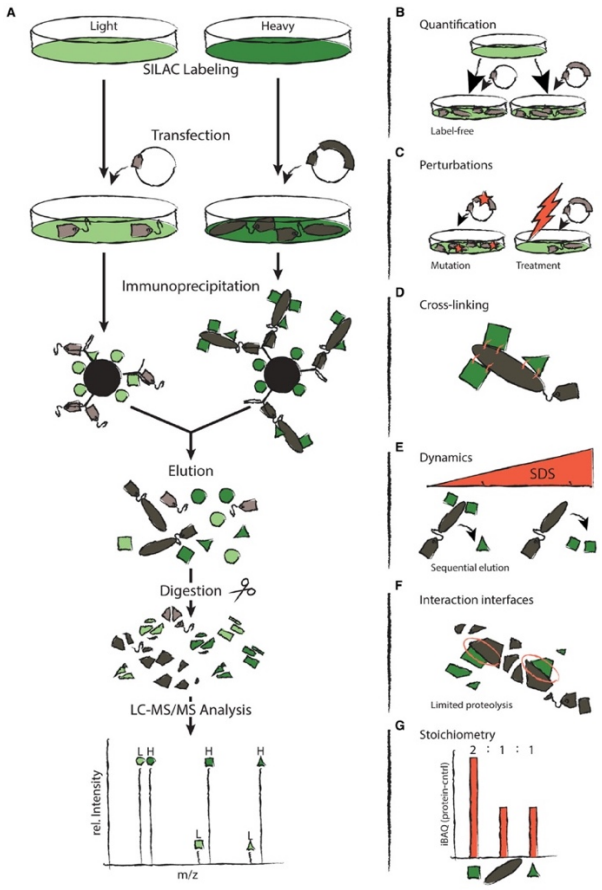


Figure 3. Overview of the stable isotope labelling with amino acids in cell culture (SILAC) procedure and its implementations in quantitative proteomic studies. (A) Cells that are differentially SILAC labelled are transfected with a tagged bait or a control vector. Immunoprecipitation is carried out using antibodies against the tag, and samples are mixed prior to elution, and subjected to enzymatic digest. The obtained peptides are then analysed through liquid-chromatography mass spectrometry (LC-MS). (**B-G**) Application of quantitative affinity purification coupled to mass spectrometry (qAP-MS) to unravel different characteristics of protein-protein interactions (PPIs). **(B)** An alternative method to SILAC is label-free quantification**. (C)** Changes in PPIs after treatments can be detected through immunoprecipitation. **(D)** When samples are cross-linked prior to extraction, transient interactions and complex structures can be studied. **(E)** Increasing concentrations of SDS during sequential elution can help to reveal PPI dynamics and submodule composition. **(F)** Interaction interfaces can be identified by using limited proteolysis. **(G)** Unravelling the stoichiometry of complexes can be carried out by comparison of the abundances of the different subunits. Figure taken from (Meyer and Selbach, 2015)

Using XIC-based proteome quantification often relies on the implementation of MRM-based methods. By coupling MRM experiments to data-dependent acquisition (DDA) in which the MRM signal triggers the acquisition of an MS/MS spectrum for a selected parent mass (Neilson *et al.*, 2011). On the one hand, such a selective approach provides MRM LFQ methods with a wider dynamic range and greater sensitivity and reproducibility than non-targeted MS methods. On the other

hand, as a balance has to be achieved between quantifying (MS) and identifying (MS/MS) a peptide, the method can only quantify a limited set of peptides in a sample. If more scanning time is invested in the identification of peptides, the number of MS quantifications decreases, resulting in a lower resolution and accuracy of XIC-based quantification. Therefor, MS^E was constructed, which implements data-independent acquisition (DIA). Here, fragmentation is carried out on all peptides, whereas ignoring their abundances. By alligning the precursor with the chromatographic elution profile of the fragmented ion, the area under the peak can be determined for the precursor ion. In plants, the usibility of $MS^{E\ i}$ has been demonstrated by the analysis of the proteome from a genetically modified soybean during risk assesment, where only expected peptides were identified (Murad and Rech, 2012).

INTERACTOMICS

Most molecular processes in the cell are carried out by proteins that form complexes via protein-protein interactions (PPIs). As such, determining a protein's localisation pattern and abundance in a cell is insufficient to understand the protein's function. In the past, a wide variety of PPI analysis methods have been constructed, of which yeast-two-hybrid (Y2H) and affinity purification coupled to MS (AP-MS) have the highest throughput (Dedecker *et al.*, 2015). Originally developed in 1989, Y2H screens assess binary interactions between two proteins through a mechanism that uses the binding of the GAL4-TF (*Saccharomyces cerevisiae*) to a reporter gene as measure of interaction (Figure 4). By genetically tagging the two proteins of interest with either the DNA-binding domain (DBD) or the trans-activation domain (TAD) of the GAL4-TF, their interactions can be identified as they will allow reconstruction of the GAL4-TF and induction of the reporter gene (Fields and Song, 1989; Dedecker *et al.*, 2015). The Y2H technique can easily be automated and has the advantage of detecting weak and transient interactions between homo- and heterodimers with dissociation constants (Kd) above 70 nM (Fang *et al.*, 2002). However, Y2H is limited to the analysis of only two proteins at the same time. Another disadvantage is the ectopic expression of proteins in a yeast background, which causes a lack of post-translational modifications. Moreover, Y2H assays suffer from relatively high false-positive rates, which range from 5-10%, due to overexpression of fusion proteins or auto-activation of transcription by other proteins. As such, Y2H screens always require other tools for validation of the identified interactions (Fang *et al.*, 2002 . Nevertheless, the technique has been successfully implemented in the construction of genome-wide PPI-maps for many organisms, including *S. cerevisiae* (Yu *et al.*, 2008), *C. elegans* (Li, 2004), *Drosophila melanogaster* (Giot, 2003), and humans (Stelzl *et al.*, 2005). Also in *Arabidopsis*, large interactome networks have been constructed (although not yet genome-wide) (de Folter, 2005, Mukhtar *et al.*, 2011).

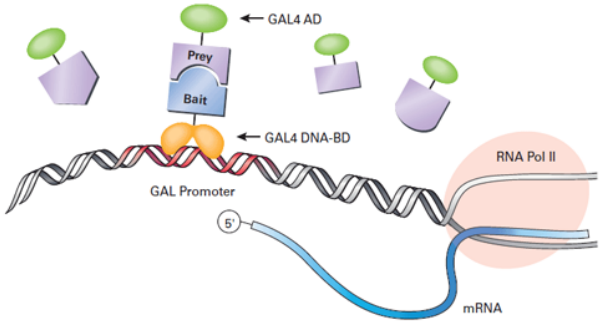


Figure 4. Schematic representation of the Yeast Two-hybrid method. A bait protein tagged with the DNA-binding domain of GAL4 will bind to the promoter of a reporter gene. If the bait interacts with the prey protein that is tagged with the transcriptional activation domain of GAL4, the reporter gene will be transcribed. Figure taken from (The yeast two-hybrid principle; clontech.com)

AP-MS, on the other hand, is more useful for the analysis of protein complexes. By tagging the protein of interest with an affinity tag, associating proteins can be co-purified with the bait, and identified through MS. Because implementation of a tag circumvents the need for protein-specific antibodies, AP-MS can theoretically be used for every clonable protein of interest in a high-throughput manner. Since its original development in yeast, where the technique was used to identify more than 1000 novel interactions that were not picked up through Y2H, a variety of AP-MS protocols and affinity tags have been developed (Dedecker *et al.*, 2015). With the recent advancements in the field of MS, mostly regarding the increasing sensitivity of the MS as well as the introduction of single-shot LC-MS, gel-based and other fractionation methods can now be avoided so that samples can be analysed in a single run. In plants, the availability of fluorescent reporter lines that contain a GFP-tagged protein as well as the availability of plant specific reproducible protocols (Wendrich *et al.*, 2017) have pushed the field towards GFP-based pull-downs (Dedecker *et al.*, 2015).

One of the trickiest parts of AP-MS-based interaction studies is the differentiation of true interactors from so-called unspecific background proteins or false positives. The latter compromise all co-purifying proteins that stick to the tag or bait during the purification process. To reduce the amount of background proteins, one of the most influential steps was the addition of a second round of purification together with the usage of an affinity tag that consists of two parts, also known as tandem affinity purification (TAP). Because of the second purification step, eluates of TAP experiments are cleaner than samples obtained from single-step AP or 'pull-down' experiments, which is advantageous for the identification of substoichiometric interactions. The first TAP tag consisted of a Calmodulin-binding peptide (CBP) and 2 IgG-binding parts of the *Staphylococcus aureus* protein A, separated from each other through a tobacco etch virus (TEV) cleavage site (Rigaut *et al.*, 1999). The tag allowed primary purification of the bait by binding of the protein A part to IgG agarose beads, after which contaminating proteins were washed away (Figure 5). Subsequent elution of complexes was then achieved

through enzymatic digest with the TEV protease, followed by a second round of purification using binding of the CBP tag to calmodulin-coated beads. After washing away the contaminating proteins for a second time, the resulting pure complexes were released from the beads through calcium chelation with EGTA, and identified through MS. Since then, a variety of TAP tags have been constructed, and TAP protocols have been optimised for implementation in a wide variety of organisms. In plants, the two most commonly used TAP tags are the TAPi tag (Rohila *et al.*, 2006) and the GS tag (Bürckstümmer *et al.*, 2006; Van Leene *et al.*, 2008). The latter combines two IgG domains of protein G with a Streptavidin-binding peptide (SBP) separated by two TEV-cleavage sites, which results in lower background levels by circumventing the calmodulin step. Thanks to some adaptations of the original protocol, complexes can now be purified from *Arabidopsis* cell suspension cultures (Van Leene *et al.*, 2007), seedlings (Van Leene *et al.*, 2014), *Medicago* roots (Goossens *et al.*, 2016), *Zea mays* (Nelissen *et al.*, 2015) and rice (Rohila *et al.*, 2006).

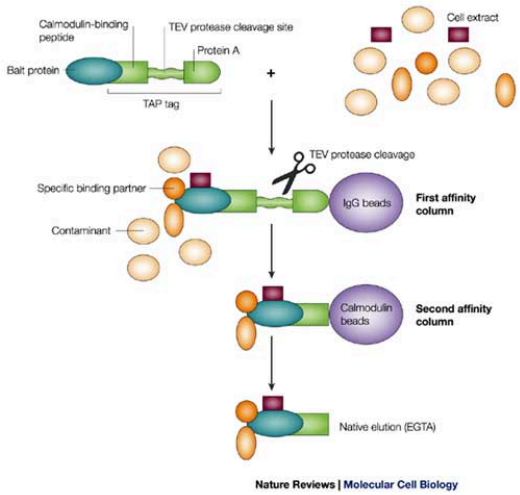


Figure 5. Schematic representation of protein complex analysis through tandem affinity purification (TAP). Cells expressing a TAP-tagged bait (where the TAP tag consists of a Protein A as immunoglobulin G (IgG)-binding domain, and a calmodulin-binding peptide separated from each other by a tobacco etch virus (TEV) protease cleavage site) are used to make a protein extract under mild conditions and subjected to TAP. Primary binding is carried out with IgG beads, whereafter the complexes can be obtained by TEV protease digest. A subsequent round of purification using calmodulin beads enhances the purity of the sample. Final elution is carried out by chelating calcium using EGTA to recover the associated complexes. Figure taken from (Huber, 2003)

To facilitate functional proteomic studies, a number of multifunctional (T)AP tags have been constructed that have a wider applicability than the standard (T)AP tags. These multifunctional tags can be used to visualise the bait and purify its interactors using different AP-based interaction studies. (Paramban *et al.*, 2004; Cheeseman and Desai, 2005; Kobayashi *et al.*, 2008; Schäffer *et al.*, 2010; Ma *et al.*, 2012; De Souza *et al.*, 2014). They provide a powerful tool for proteomic research

by combining fluorescent protein tags for visualisation (Giepmans, 2006; Voss *et al.*, 2013) with other single or tandem AP tags. Their applicability across species, such as yeast, mammalian cells, *C. elegans, E. coli,* tobacco leaf extracts and *Aspergillus nidulans* has been reported (Paramban *et al.*, 2004; Cheeseman and Desai, 2005; Kobayashi *et al.*, 2008; Schäffer *et al.*, 2010; Ma *et al.*, 2012; De Souza *et al.*, 2014). For example, implementation of the LAP tag has helped to visualise TDP-43 (a protein linked with the ALS disease) complexes and to purify its complexes via quantitative TAP-MS. This resulted in the idenfification of interactions between TDP-43 and heterogeneous nuclear ribonucleoproteins family proteins, and also with components of Drosha microprocessor complexes, consistent with roles for TDP-43 in both mRNA processing and microRNA biogenesis (Ling *et al.*, 2010). In the plant field, successful implementation of these multifunctional tags for functional protein analysis is still missing.

Table 1. Overview of multifunctional tags and their applicability.

Tag Name	Parts	Applications	Biological Background	Original Reference
Dual functional GFP tag	GFP-2xTEV-6x His	Localisation and TAP	*E. coli, N. benthamiana* leaf extracts	(Paramban *et al.*, 2004)
LAP tag	(C): 6x His-($3C_{rhino}$)-GFP (N): GFP-(TEV)-S-tag	Localisation and TAP	Metazoans	(Cheeseman & Desai, 2005)
mfGFP	8x His-SBP-cMyc in GFP(/mCherry) loop	Immunofluorescense, EM and TAP	Mammalians	(Kobayashi *et al.*, 2008)
SnAvi tag	Avi tag-(TEV)-EGFP- anti-SB1 epitope	Localisation and TAP	*C. elegans*	(Schäffer *et al.*, 2010)
MAP tag	His_8-SBP-FLAG epitope in mVENUS loop	Localisation and TAP	Mammalians, *C. elegans* and *S. pombe*	(Ma *et al.*, 2012ce)
DLAP tag	10AA linker-GFP-S-tag	Localisation and AP	*A. nidulans*	(DeSouza *et al.*, 2014)

(C): C-terminal tag; (N): N-terminal tag. Fluorescent molecules are underlined.

The second method to identify true interactors from background is based on the availability ofM quantitative information on the proteins that are identified with AP-MS. Methods that can be used for protein quantification are already discussed above (see Proteomics). By integrating such quantitative data, *bona fide* interactors can more easily be distinguished, as they are expected to be enriched after purification in comparison to a negative control (Dedecker *et al.*, 2015). With the constant increase in sensitivity of MS, the field of AP-MS is now even shifiting more towards affinity enrichment (AE) rahter than purification. The idea behind it is that the higher resolution of the MS allows the usage of endogenous promoters as well as the implemention of protocols that are less stringent (favouring weaker interactions) and create a higher background. As the intenstity of background proteins is linked to their abundances in the sample as well as to their affinity to the beads, a "beadome" that represents a scaled downversion of the proteome can therefore be constructed (Keilhauer *et al.*, 2014). The reproducibility of these background binders is of course linked to their intensities, but remains rather

constant between experiments. This implies that both negative controls or other pull-down experiments can be used to idenify bait-specific enriched proteins. It also shows that the unspecific binders can be used as a quality control, as deviations of the anticipated behaviour can indicate lower qualities of samples (Keilhauer *et al.*, 2014). Large sale studies that use such an AE-MS approach to study the interactome using endogenous promoters have been published for yeast (Keilhauer *et al.*, 2014) and humans (Hein *et al.*, 2015), but not yet for plants.

Besides using the quantitative data to identify interactors of a bait, the field of quantitative proteomics is now also starting to look into the abundances of the interactors and the stoichiomtery of their interactions via computational methods. The idea is that when the protein is expressed from its endogenous promoter, interaction stoichiometries can reflect the stability of a PPI and will depend on the extent of their interaction. However, the abundance of the interactor in the cell also has to be taken into account, as this might influence the recoverability of the interaction after AP. As such, the quantitative data is now also being used to establish the relative cellular abundance of interactors at the same time. The usability of such an approach was recently published for the human interactome (Hein *et al.*, 2015). Interestingly, this study described how this new knowledge on the obeserved interactions can be used to explain phenotypes and genetic associations.

Although the AE approach is best for experimental conditions in which the controls are dirty, as is the case for pull-down samples, quantitative (q)TAP-MS has been successfully applied. In human cells we already gave the example of the TDP-43 protein (see above). In plants, qTAP-MS revealed dynamic changes in complex composition of the maize ANGUSTIFOLIA3 (ZmAN3) complex in different developmental tissues of the leaf (Nelissen *et al.*, 2015). With the constantly increasing sensitivity of the mass spectrometers, qAP-MS is expected to become one of the most powerful tools to analyse PPIs. There are, however, some pitfalls of the technology. The first one is the presence of the tag, which might interfere with the functionality of the bait. Combined with overexpression of the bait, which is often required to out-compete the protein's endogenous counterpart, this can lead to a lot of false-positive or false-negative interactions. An alternative is the expression of the fusion protein from its endogenous sequences in a mutant background, which also allows immediate validation of full protein functionality through complementation analysis. However, not every mutation leads to a complementable phenotype, and complementation is not always indicative of full protein functionality (Zhou *et al.*, 2011). Also, our current knowledge on enhancer and promoter sequences that control gene expression is still limited for transgene production. Ideally, the endogenous DNA is tagged directly through homologous recombination, which is not yet feasible in plants but might become possible in the future thanks to advances in the CRISPR/Cas9 genome editing.

Besides interacting with other proteins, some proteins, *i.e.*, transcription factors, also bind to DNA in the nucleus, thereby influencing the expression of the targeted genes. One of the most commonly used techniques to identify target

genes of a TF is Chromatin Immunoprecipitation (ChIP). To avoid the need for protein specific antibodies, a tag, such as GFP or the hemagglutinin (HA) tag, is often added to the protein of interest. ChIP allows genome-wide screening for binding sites by cross-linking the tagged bait to the DNA with, for example, formaldehyde (Johnson *et al.*, 2007). The cross-linked chromatin is then sheared into smaller fragments, and subjected to affinity purification against the bait to enrich bound DNA regions. These regions can then be identified through hybridisation with a DNA microarray (ChIP-chip) (Massie and Mills, 2008) or via NGS (ChIP-seq) (Ho *et al.*, 2011) (Figure 6).

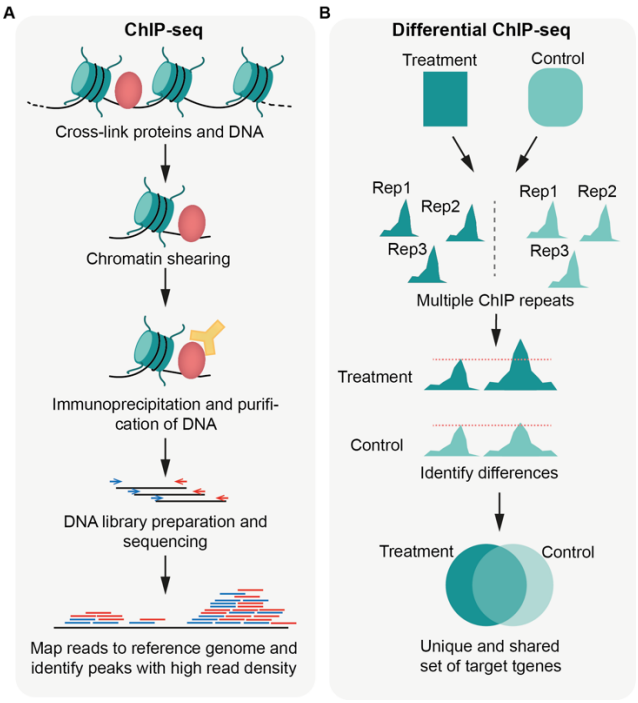


Figure 6. Identifying the target genes of a transcription factor through Chromatin Immunoprecipitation (ChIP) coupled to next-generation sequencing. (A) Schematic overview of the main steps in the ChIP-seq protocol. Cross-linked chromatin is sheared through sonication or via enzymatic digestion. Bound DNA is precipitated via an antibody against the target protein, and recovered via reverse cross-linking. The enriched DNA can then be identified through sequencing and mapping to the reference genome. **(B)** Differential ChIP-seq analysis allows identification of differences in binding affinity of target genes between two samples by quantifying changes in the number of reads obtained per peak.

Because the search space of ChIP-chip is limited to whatever DNA is bound on the microarray and NGS methods are becoming cheaper and more accurate, ChIP-seq is now becoming the new standard (Ho *et al.*, 2011; Ricardi *et al.*, 2014; Pautler *et al.*, 2015; Yang *et al.*, 2016). To identify the protein's binding sites, however, ChIP-seq requires mapping of the sequenced reads to a reference genome and downstream computational analysis. A common method in the identification of

enriched peaks is Model-based analysis of ChIP-seq (MACS) (Zhang *et al.*, 2008). ChIP-seq can also be applied to identify changes in binding affinity of a TF in different developmental contexts or conditions, a technique better known as differential ChIP-seq (Steinhauser *et al.*, 2016). Statistical tests that are commonly used to assess such differential binding include DiffBind, csaw and THOR (Steinhauser *et al.*, 2016; Lun and Smyth, 2016; Allhoff *et al.*, 2016). It is important to mention that the choice of downstream analysis tools can have a great influence on the final result of the experiment.

Another interaction that is becoming more and more of interest is the interaction of proteins with RNA. There are two commonly used tools that allow the identification of such binding events. Similar to ChIP, cross-linking immunoprecipitation (CLIP) uses cross-linking of proteins to the RNA via UV light (Meng *et al.*, 2014). The resulting bound fragments are then digested with nucleases and analysed via sequencing. Another option, RIP (RNA immunoprecipitation), couples immunoprecipitation to reverse transcription followed by sequencing (Terzi and Simpson, 2009).

METABOLOMICS

The final field of study that is indispensable in plant systems biology is metabolomics. Metabolites are generally small molecules, such as sugars, amino acids or fatty acids, which are chemically metabolised by enzymes. Often they play a key role in many biological processes, and form the chemical basis of crop yield and quantity (Hall *et al.*, 2007). There are two types of metabolites: primary metabolites, *i.e.*, molecules that are necessary for plant growth and development, and secondary metabolites, which are not essential for survival but are crucial for plant responses to stresses (Hong *et al.*, 2016). Given their function, the structure and abundance of primary metabolites are highly conserved across plant kingdoms, whereas secondary metabolites are more heterogenic (Scossa *et al.*, 2016). This diversity in biochemical nature makes it rather difficult to study metabolites, as their structures and abundances vary greatly. Therefore, metabolomic studies often combine several techniques to analyse complex samples in large-scale set-ups (Hegeman, 2010). The most commonly applied techniques to analyse the metabolome of plants are non-destructive NMR (nuclear magnetic resonance spectroscopy) and MS-based methods including LC–MS (liquid chromatography–MS), GC-MS (gas chromatography-MS), CE-MS (capillary electrophoresis–MS), and FI-ICR–MS (Fourier transform ion cyclotron resonance–MS)(Okazaki and Saito, 2011; Khakimov *et al.*, 2014; Hong *et al.*, 2016). NMR results are highly reproducible, and have the advantage that samples are easy to prepare. MS-based techniques, on the other hand, have a higher sensitivity than NMR, allowing detection of most cellular metabolites (Bunnik and Le Roch, 2013). With the introduction of high-resolution sampling methods, metabolomics studies can now even be carried out at the subcellular level as well as in single cells (Moussaieff *et al.*, 2013; Sweetlove *et al.*, 2014; Misra *et al.*, 2014). The power of such techniques for plant metabolic studies has been demonstrated in a wide variety of applications for many common

plant species and crops, including rice, tomato, wheat and maize (Rao *et al.*, 2014; Hu *et al.*, 2014; Bénard *et al.*, 2015). However, because both MS- and NMR-based techniques have their limitations, metabolomics analyses are now more and more switching to combined approaches.

In contrast to proteomic and transcriptomic studies, the extent of the metabolome of a plant cannot be predicted based on the genome. Hence, by taking a snapshot of the metabolome, a large quantity of data obtained via accurate and high-throughput peak annotation is combined to identify new genes or pathways (Tohge *et al.*, 2014). Additionally, the regulatory networks between all the metabolites in plants appear to be very complicated. Therefore, in order to gain an insight into the relationship between metabolic output and genotype, metabolic studies focus on the key components in these regulatory networks and their interactions. Combined with transcriptomics analysis, such metabolomics approaches have been successfully implemented to unravel the regulation of metabolic fluxes and the concentrations of these small molecules in plants (Kleessen and Nikoloski, 2012; Toubiana *et al.*, 2013).

Metabolomic tools are also useful to obtain a deeper understanding of the processes underlying plant development. Screening crops for metabolite quantitative trait loci (mQTL) and metabolite genome-wide association studies (mGWAS) can help to discover genes that are linked to a metabolic phenotype in genetic variants, and results in much more information than classic macroscopical phenomic approaches. As they provide an insight into plant biology from a metabolic point of view, both mGWAS and mQTL research studies have been conducted in a wide variety of plants, including *Arabidopsis,* tomato, wheat, rice and rape (Schauer *et al.*, 2006; Yang *et al.*, 2009; Matsuda *et al.*, 2012; Vercruyssen *et al.*, 2014; Hill *et al.*, 2015). Combining reverse genetics approaches, such as RNAi and knock-out mutants (Bringaud *et al.*, 2015) and overexpressors, with metabolomics tools can also be used to explore the function of enzymes in metabolic pathways, and can as such aid the functional annotation of the causing genes (Figure 7). This was demonstrated in *Arabidopsis* with the characterisation of phenylalanine ammonia-lyase (Cravatt, 2004). Similar to the other functional genomics fields, the introduction of CRISPR/Cas9-based methods for gene targeting will greatly enhance our comprehension of plant metabolic processes, as it allows the simultaneous analysis several genes (Xing *et al.*, 2014). As such, screening the metabolic profile of multiple gene mutants will provide an insight into the interaction of two or more genes in a given metabolic pathway. Combined with mGWAS and mQTL analysis, our understanding of the genetic and biochemical mechanisms that drive metabolic pathways in plants will become more profound. On the other hand, integrating data from transcriptomics and proteomic studies with metabolic profiling can also help to broaden our knowledge on plant development and responses to environmental stresses.

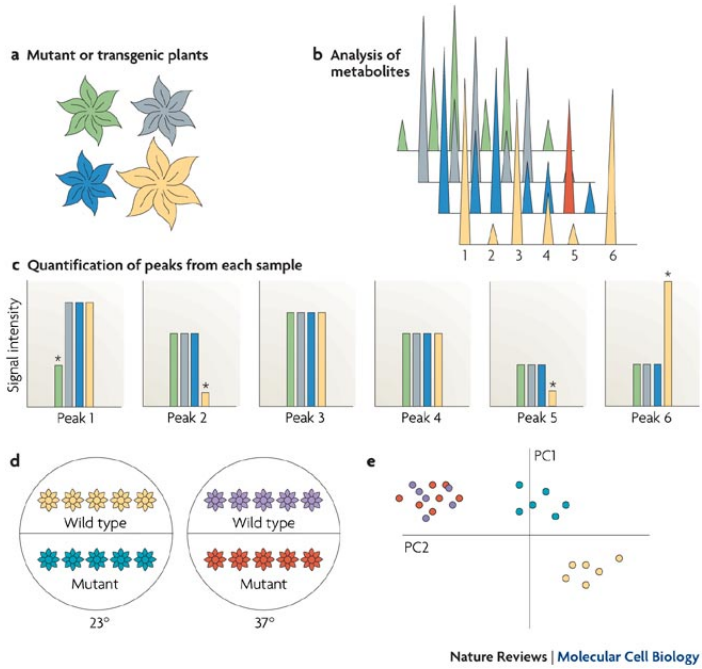


Figure 7. Implementing metabolomics and reverse genetics approaches for the analysis of the effect of genetic or environmental perturbations in plants. Analysis of small metabolites can help to identify differences in phenotypes amongst genetic variants in natural populations or in variants obtained through mutagenesis or transformation **(A-C).** Although the phenotype (both colour and size) of the blue and yellow plants in panel **A** are distinct, it is difficult to infer any information on their metabolic pathways based on morphology only. Panel **B** represents a schematic view of six peaks (using the same colour scheme as the plants in **A**) with their respective intensities as identified by gas or liquid chromatography coupled to quantitative mass spectrometry. Additionally, a red peak is present among the peaks identified in the blue plant which represents a new metabolite that is unique to this mutant. Downstream quantitative comparison of the remaining six peaks is shown in panel **C**. Asterisk indicates statistically significant differences between plants. Metabolomics and metabolite profiling can also be implemented in the evaluation of plant responses to environmental signals, shown in panels **D** and **E**. Here, two growth conditions (23°C versus 37°C) were tested on wild-type and high-temperature-tolerant plants. Metabolic profiling showed that wild-type plants grown at 23°C are phenotypically distinct from the other samples, whereas high-temperature-tolerant mutants showed high similarities to the wild-type plants grown at 37°C. PC: principal component. Figure taken from (Last *et al.*, 2007)

In conclusion, combining data from genomics, transcriptomics, proteomics, metabolomics and interactomics experiments is a very powerful approach to increase our understanding of the function of single genes in biological processes, such as development, growth and responses to the environment. By providing system-wide analysis tools, each subfield covers a different type of information and aids in annotating the complex genome of plants. Moreover, with the rapid improvement of MS- and NGS-based techniques and the introduction of CRISPR/Cas9 gene targeting methods, functional annotation of unknown genes will become even more straightforward.

REFERENCES

Abdeeva I, Abdeev R, Bruskin S, Piruzi E. 2012. Transgenic Plants as a Tool for Plant Functional Genomics.Ozden ifti Y ed. Transgenic Plants - Advances and Limitations. InTech.

Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy ASN. 2016. A survey of the sorghum transcriptome using single-molecule long reads. Nature Communications **7**, 11706.

Aebersold R, Mann M. 2016. Mass-spectrometric exploration of proteome structure and function. Nature **537**, 347–355.

Aldaz CM. 2003. Serial Analysis of Gene Expression (SAGE) in Cancer Research. Expression Profiling of Human Tumors. New Jersey: Humana Press, 47–60.

Allhoff M, Seré K, Pires JF, Zenke M. 2016. Differential peak calling of ChIP-seq signals with replicates with THOR. Nucleic Acids Research **44**, e153

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR. 2003. Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science **301**, 653–657.

Alonso JM, Ecker JR. 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nature Reviews Genetics **7**, 524–536.

Amanchy R, Kalume DE, Pandey A. 2005. Stable Isotope Labelling with Amino Acids in Cell Culture (SILAC) for Studying Dynamics of Protein Abundance and Posttranslational Modifications. Science Signaling **267** pl2.

Ben-Amar A, Daldoul S, Reustle GM, Krczal G, Mliki A. 2016. Reverse Genetics and High Throughput Sequencing Methodologies for Plant Functional Genomics. Current genomics **17**, 460–475.

Bénard C, Bernillon S, Biais B, Osorio S, Maucourt M, Ballias P, Deborde C, Colombié S, Cabasson C, Jacob D, Vercambre G, Gautier H, Rollin D, Génard M, Fernie AR, Gibon Y, Moing A. 2015. Metabolomic profiling in tomato reveals diel compositional changes in fruit affected by source–sink relationships. J Exp Bot **66**, 3391–3404.

Bolger ME, Arsova B, Usadel B. 2017. Plant genome and transcriptome annotations: from misconceptions to simple solutions. Brief Bioinform, bbw135.

Bringaud F, Biran M, Millerioux Y, Wargnies M, Allmann S, Mazet M. 2015. Combining reverse genetics and nuclear magnetic resonance-based metabolomics unravels trypanosome-specific metabolic pathways. Molecular Microbiology **96**, 917–926.

Bunnik EM, Le Roch KG. 2013. An Introduction to Functional Genomics and Systems Biology. Advances in Wound Care **2**, 490–498.

Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superti-Furga G, Bauch A. 2006. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat Methods **3**, 1013–1019.

Chan Y-K. 2009. Breeding Papaya (Carica papaya L.).Jain SM, Priyadarshan PM eds. Breeding Plantation Tree Crops: Tropical Species. New York, NY: Springer New York, 121–159.

Cheeseman IM, Desai A. 2005. A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans. Science Signaling **2005**, pl1–pl1.

Collings DA. 2013. Subcellular localization of transiently expressed fluorescent fusion proteins. Methods Mol Biol **1069**, 227–258.

Consortium TRF-LC, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team, Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T; Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K; RIKEN, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y, Yasunishi A. 2003. Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from japonica Rice. Science **301**, 376–379.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology **26**,

1367–1372.

Cravatt B. 2004. Faculty of 1000 evaluation for Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism.

de Folter S, Immink RGH, Kieffer M,Pařenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM,Davies B, Angenent GC. 2005. Comprehensive Interaction Map of the *Arabidopsis* MADS Box Transcription Factors. The Plant Cell **17**, 1424–1433.

De Souza CP, Hashmi SB, Osmani AH, Osmani SA. 2014. Application of a New Dual Localization-Affinity Purification Tag Reveals Novel Aspects of Protein Kinase Biology in Aspergillus nidulans (J-H Yu, Ed.). PLoS ONE **9**, e90911.

Dedecker M, Van Leene J, De Jaeger G. 2015. Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry. Current Opinion in Plant Biology **24**, 1–9.

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN. 2008. Cell Identity Mediates the Response of *Arabidopsis* Roots to Abiotic Stress. Science **320**, 942–945.

Evans C, Noirel J, Ow SY, Salim M, Pereira-Medrano AG, Couto N, Pandhal J, Smith D, Pham TK, Karunakaran E, Zou X, Biggs CA, Wright PC. 2012. An insight into iTRAQ: where do we stand now? Analytical and Bioanalytical Chemistry **404**, 1011–1027.

Fang Y, Macool DJ, Xue Z, Heppard EP, Hainey CF, Tingey SV, Miao G-H. 2002. Development of a high-throughput yeast two-hybrid screening system to study protein-protein interactions in plants. Molecular Genetics and Genomics **267**, 142–153.

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

Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. 2010. Super-SILAC mix for quantitative proteomics of human tumor tissue. Nature Methods **7**, 383–385.

Gichner T. 1995. Meyerowitz E.M., C.R. Somerville (ed.):Arabidopsis. Biologia Plantarum **37**, 540–540.

Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD. 2008. Cell-specific nitrogen responses mediate developmental plasticity. Proceedings of the National Academy of Sciences **105**, 803–808.

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. 2003. A protein interaction map of Drosophila melanogaster. Science **302**, 1727–1736.

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. The Plant Journal **88**, 476–489.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y. 2013. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nature Genetics **45**, 51–58.

Hall RD, Brouwer ID, Fitzgerald MA. 2008. Plant metabolomics and its potential application for human nutrition. Physiologia Plantarum **132,** 162-175

Handakumbura PP, Hixson KK, Purvine SO, Jansson C, Paša Tolić L. 2016. *Plant iTRAQ-Based Proteomics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics **107**, 1–8.

Hegeman AD. 2010. Plant metabolomics - meeting the analytical challenges of comprehensive metabolite analysis. Briefings in Functional Genomics **9**, 139–148.

Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, Hyman AA, Mann M. 2015. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. Cell **163**, 712–723.

Heller MJ. 2002. DNA microarray technology: devices, systems, and applications. Annual review of

biomedical engineering **4**, 129–153.
Hill CB, Taylor JD, Edwards J, Mather D, Langridge P, Bacic A, Roessner U. 2015. Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. Plant Science **233**, 143–154.
Ho JWK, Bishop E, Karchenko PV, Nègre N, White KP, Park PJ. 2011. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC Genomics **12**, 134.
Hong J, Yang L, Zhang D, Shi J. 2016. Plant Metabolomics: An Indispensable System Biology Tool for Plant Science. International Journal of Molecular Sciences **17**, 767.
Hu C, Shi J, Quan S, Cui B, Kleessen S, Nikoloski Z, Tohge T, Alexander D, Guo L, Lin H, Wang J, Cui X, Rao J, Luo Q, Zhao X, Fernie AR, Zhang D. 2014. Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. Scientific Reports **4**.
Huber LA. 2003. Is proteomics heading in the wrong direction? Nature Reviews Molecular Cell Biology **4**, 74–80.
Ichikawa T, Nakazawa M, Kawashima M, Iizumi H, Kuroda H, Kondou Y, Tsuhara Y, Suzuki K, Ishikawa A, Seki M, Fujita M, Motohashi R, Nagata N, Takagi T, Shinozaki K, Matsui M. 2006. The FOX hunting system: an alternative gain-of-function gene hunting technique. The Plant Journal **48**, 974–985.
Inna Abdeeva RASB, Piruzian E. 2012. Transgenic Plants as a Tool for Plant Functional Genomics. Transgenic Plants - Advances and Limitations. InTech.
Ito T, Meyerowitz EM. 2000. Overexpression of a Gene Encoding a Cytochrome P450, CYP78A9, Induces Large and Seedless Fruit in *Arabidopsis*. The Plant Cell Online **12**, 1541–1550.
Itzhak DN, Tyanova S, Cox J, Borner GH. 2016. Global, quantitative and dynamic mapping of protein subcellular localization. eLife **5**, e16950.
Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnology **15**, 16.
Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P; French-Italian Public Consortium for Grapevine Genome Characterization. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463–467.
Ji W, Cong R, Li S, Li R, Qin Z, Li Y, Zhou X, Chen S, Li J. 2016. Comparative Proteomic Analysis of Soybean Leaves and Roots by iTRAQ Provides Insights into Response Mechanisms to Short-Term Salt Stress. Frontiers in Plant Science **7**, 573.
Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science **316**, 1497–1502.
Karimi M, Bleys A, Vanderhaeghen R, Hilson P. 2007. Building blocks for plant gene assembly. Plant Physiology **145**, 1183–1191.
Keilhauer EC, Hein MY, Mann M. 2014. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). Molecular & Cellular Proteomics **14**, 120–135.
Keller A, Meese E. 2014. *Nucleic acids as molecular diagnostics*.
Khakimov B, Bak S, Engelsen SB. 2014. High-throughput cereal metabolomics: Current analytical technologies, challenges and perspectives. Journal of Cereal Science **59**, 393–418.
Kikuchi S. 2003. Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from japonica Rice. **301**, 376–379.**Kleessen S, Nikoloski Z**. 2012. Dynamic regulatory on/off minimization for biological systems under internal temporal perturbations. BMC Systems Biology **6**, 16.
Kleessen S, Nikoloski Z. 2012. Dynamic regulatory on/off minimization for biological systems under internal temporal perturbations. BMC Systems Biology **6**, 16.
Kobayashi T, Morone N, Kashiyama T, Oyamada H, Kurebayashi N, Murayama T. 2008. Engineering a Novel Multifunctional Green Fluorescent Protein Tag for a Wide Variety of Protein Research (A Imhof, Ed.). PLoS ONE **3**, e3822.
Köcher T, Pichler P, Schutzbier M, Stingl C, Kaul A, Teucher N, Hasenfuss G, Penninger JM, Mechtler K. 2009. High Precision Quantitative Proteomics Using iTRAQ on an LTQ Orbitrap: A New Mass Spectrometric Method Combining the Benefits of All. Journal of Proteome Research **8**, 4743–4752.
Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fässler R,

Mann M. 2008. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell **134**, 353–364.

Last RL, Jones AD, Shachar-Hill Y. 2007. Towards the plant metabolome and beyond. Nature Reviews Molecular Cell Biology **8**, 167–174.

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, Wohlschlegel JA, Jacobsen SE. 2010. A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in *Arabidopsis*. Current biology **20**, 951–956.

Lewandowska D, Have Ten S, Hodge K, Tillemans V, Lamond AI, Brown JWS. 2013. Plant SILAC: stable-isotope labelling with amino acids of arabidopsis seedlings for quantitative proteomics. PLoS ONE **8**, e72207.

Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. 2004. A map of the interactome network of the metazoan *C. elegans*. Science **303**, 540–543.

Lim G-H, Shine MB, de Lorenzo L, Yu K, Cui W, Navarre D, Hunt AG, Lee J-Y, Kachroo A, Kachroo P. 2016. Plasmodesmata Localizing Proteins Regulate Transport and Signaling during Systemic Acquired Immunity in Plants. Cell host & microbe **19**, 541–549.

Ling S-C, Albuquerque CP, Han JS, Lagier-Tourenne C, Tokunaga S, Zhou H, Cleveland DW. 2010. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proceedings of the National Academy of Sciences of the United States of America **107**, 13318–13323.

Lisec J, Steinfath M, Meyer RC, Selbig J, Melchinger AE, Willmitzer L, Altmann T. 2009. Identification of heterotic metabolite QTL in *Arabidopsis thaliana* RIL and IL populations. The Plant Journal **59**, 777–788

Lun ATL, Smyth GK. 2016. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Research **44**, e45–e45.

M Perez-de-Castro A, Vilanova S, Cañizares J, Pascual L, M Blanca J, J Diez M, Prohens J, Pico B. 2012. Application of Genomic Tools in Plant Breeding. Current genomics **13**, 179–195.

Ma H, McLean JR, Chao LF-I, Mana-Capelli S, Paramasivam M, Hagstrom KA, Gould KL, McCollum D. 2012. A Highly Efficient Multifunctional Tandem Affinity Purification Approach Applicable to Diverse Organisms. Molecular & Cellular Proteomics **11**, 501–511.

Martínez-Esteso MJ, Casado-Vela J, Sellés-Marchart S, Pedreño MA, Bru-Martínez R. 2014. Differential plant proteome analysis by isobaric tags for relative and absolute quantitation (iTRAQ). Methods in molecular biology (Clifton, N.J.) **1072**, 155–169.

Massie CE, Mills IG. 2008. ChIPping away at gene regulation. EMBO reports **9**, 337–343.

Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, Yonemaru J-I, Ebana K, Yano M, Saito K. 2012. Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. The Plant Journal **70**, 624–636.

McCallum CM, Comai L, Greene EA, Henikoff S. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiology **123**, 439–442.

Mehboob-ur-Rahman, Rahmat Z, Gul M, Zafar Y. 2016. Plant functional genomics: Approaches and applications. The Next Generation of Genetic Engineering. Applied Molecular Biotechnology. CRC Press, 157–186.

Meng J, Shi L, Luan Y. 2014. Plant microRNA-Target Interaction Identification Model Based on the Integration of Prediction Tools and Support Vector Machine. PLoS ONE **9**, e103181.

Meyer K, Selbach M. 2015. Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. Frontiers in Genetics **6**, 237.

Meyerowitz EM, Somerville CR. 1994. *Arabidopsis*. Cold Spring Harbor monograph series (USA).

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M. 2008. The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature

452, 991-996

Misra BB, Assmann SM, Chen S. 2014. Plant single-cell and single-cell-type metabolomics. Trends in Plant Science **19**, 637–646.

Mittler R, Feng X, Cohen M. 1998. Post-Transcriptional Suppression of Cytosolic Ascorbate Peroxidase Expression during Pathogen-Induced Programmed Cell Death in Tobacco. The Plant Cell Online **10**, 461–474.

Moussaieff A, Rogachev I, Brodsky L, Malitsky S, Toal TW, Belcher H, Yativ M, Brady SM, Benfey PN, Aharoni A. 2013. High-resolution metabolic mapping of cell types in plant roots. Proceedings of the National Academy of Sciences **110**, E1232–E1241.

Murad AM, Rech EL. 2012. NanoUPLC-MSE proteomic data assessment of soybean seeds using the Uniprot database. BMC Biotechnology **12**, 82.

Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M. 2012. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Molecular & Cellular Proteomics **11**, M111.013722.

Nakamura H, Hakata M, Amano K, Miyao A, Toki N, Kajikawa M, Pang J, Higashi N, Ando S, Toki S, Fujita M, Enju A, Seki M, Nakazawa M, Ichikawa T, Shinozaki K, Matsui M, Nagamura Y, Hirochika H, Ichikawa H.. 2007. A genome-wide gain-of-function analysis of rice genes using the FOX-hunting system. Plant Molecular Biology **65**, 357–371.

Nakazawa M, Ichikawa T, Ishikawa A, Kobayashi H, Tsuhara Y, Kawashima M, Suzuki K, Muto S, Matsui M. 2003. Activation tagging, a novel tool to dissect the functions of a gene family. The Plant Journal **34**, 741–750.

Nakazawa M, Yabe N, Ichikawa T, Yamamoto YY, Yoshizumi T, Hasunuma K, Matsui M. 2008. DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. The Plant Journal **25**, 213–221.

Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA. 2011. Less label, more free: approaches in label-free quantitative mass spectrometry. (MJ Dunn, Ed.). Proteomics **11**, 535–553.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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.

Okazaki Y, Saito K. 2011. Recent advances of metabolomics in plant biotechnology. Plant Biotechnology Reports **6**, 1–15.

Pan L, Shah AN, Phelps IG, Doherty D, Johnson EA, Moens CB. 2015. Rapid identification and recovery of ENU-induced mutations with next-generation sequencing and Paired-End Low-Error analysis. BMC Genomics **16**, 83.

Paramban RI, Bugos RC, Wen Su W. 2004. Engineering green fluorescent protein as a dual functional tag. Biotechnology and Bioengineering **86**, 687–697.

Paterson AH, Bowers J, Bruggmann R, Dubchak I, Grimwood J, Gunflach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus A, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, KresovichS, McCann MC, Ming RF, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature **457**, 551–556.

Pautler M, Eveland AL, LaRue T, Yang F, Weeks R, Lunde C, Je BI, Meeley R, Komatsu M, Vollbrecht E, Sakai H, Jackson D. 2015. FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize. The Plant Cell **27**, 104–120.

Qureshi IA, Mehler MF. 2011. Advances in epigenetics and epigenomics for neurodegenerative diseases. Current neurology and neuroscience reports **11**, 464–473.

Rao J, Cheng F, Hu C, Rao J, Cheng F, Hu C, Quan S, Lin H, Wang J, Chen G, Zhao X, Alexander D, Guo L, Wang G, Lai J, Zhang D, Shi J. 2014. Metabolic map of mature maize kernels. Metabolomics **10**, 775–787.

Ricardi MM, González RM, Zhong S, Domínguez PG, Duffy T, Turjanski PG, Salgado Salter JD, Alleva K, Carrari F, Giovannoni JJ, Estévez JM, Iusem ND. 2014. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought-stress-responsive transcription factor. BMC Plant Biology **14**, 29.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology **17**, 1030–1032.

Rinaldo AR and Ayliffe M. 2015. Gene targeting and editing in crop plants: a new era of precision opportunities. Molecul Breeding **35:**40

Rohde A, Morreel K, Ralph J, Goeminne G. 2004. Molecular phenotyping of the pal1 and pal2 mutants of *Arabidopsis* thaliana reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate The Plant Cell **16,** 2749–2771

Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, Zhu JK, Ronald P, Fromm ME. 2006. Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. The Plant Journal **46**, 1–13.

Ronald PC. 2014. Lab to Farm: Applying Research on Plant Genetics and Genomics to Crop Improvement. PLoS Biology **12**, e1001878.

Sakurai T, Plata G, Rodríguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M. 2007. Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. BMC Plant Biology **7**, 66.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology.

Sasaki T, Burr B. 2000. International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Current Opinion in Plant Biology **3**, 138–141.

Schäffer U, Schlosser A, Müller KM, Schäfer A, Katava N, Baumeister R, Schulze E. 2010. SnAvi – a new tandem tag for high-affinity protein-complex purification. Nucleic Acids Research **38**, e91–e91.

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-MElis A, Bruedigam C, Kopka J, Willmitzer L, Zamier D, Fernie AR. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology **24**, 447–454.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquit E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacy G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. Nature **463**, 178–183.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science **326**, 1112–1115.

Scossa F, Brotman Y, de Abreu e Lima F, Willmitzer L, Nikoloski Z, Tohge T, Fernie AR. 2016. Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. Plant science: an international journal of experimental plant biology **242**, 47–64.

Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE **8**, e61005.

Senthil-Kumar M, Hema R, Suryachandra TR, Ramegowda HV, Gopalakrishna R, Rama N, Udayakumar M, Mysore KS. 2010. Functional characterization of three water deficit stress-induced genes in tobacco and *Arabidopsis*: an approach based on gene down regulation. Plant physiology and biochemistry : PPB **48**, 35–44.

Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR-Cas9. Nature Reviews Genetics **16**, 299–311.

Shen L, Hua Y, Fu Y, Li J, Liu Q, Jiao X, Xin G, Wang J, Wang X, Yan C, Wang K. 2017. Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. Science China. Life sciences **309**, 1–10.

Shi J, Chen Y, Xu Y, Ji D, Chen C, Xie C. 2017. Differential Proteomic Analysis by iTRAQ Reveals the Mechanism of Pyropia haitanensis Responding to High Temperature Stress. Scientific Reports **7**, 44734.

Shiio Y, Aebersold R. 2006. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. Nature Protocols **1**, 139–145.

Steinhauser S, Kurzawa N, Eils R, Herrmann C. 2016. A comprehensive comparison of tools for differential ChIP-seq analysis. Brief Bioinform **17**, 953–966.

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. 2005. A human protein-protein interaction network: a resource for annotating the proteome. Cell **122**, 957–968.

Sweetlove LJ, Obata T, Fernie AR. 2014. Systems analysis of metabolic phenotypes: what have we learnt? Trends in Plant Science **19**, 222–230.

Terzi LC, Simpson GG. 2009. *Arabidopsis* RNA immunoprecipitation. The Plant Journal **59**, 163–168.

Tohge T, de Souza LP, Fernie AR. 2014. Genome-enabled plant metabolomics. Journal of Chromatography B **966**, 7–20.

Toubiana D, Fernie AR, Nikoloski Z, Fait A. 2013. Network analysis: tackling complex data to study plant metabolism. Trends in Biotechnology **31**, 29–36.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D.I. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313**, 1596–1604.

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. Nature Protocols **10**, 169–187.

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, Inzé D, Witters E, De Jaeger G. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis* thaliana. Molecular & Cellular Proteomics **6**, 1226–1238.

Van Leene J, Witters E, Inzé D, De Jaeger G. 2008. Boosting tandem affinity purification of plant protein complexes. Trends in Plant Science **13**, 517–520.

van Ruissen F, Ruijter JM, Schaaf GJ, Asgharnegad L, Zwijnenburg DA, Kool M, Baas F. 2005. Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. BMC Genomics **6**, 91.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L, Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. Plant Cell **26**, 210–229.

Voss U, Larrieu A, Wells DM. 2013. From jellyfish to biosensors: the use of fluorescent proteins in plants. The International journal of developmental biology **57**, 525–533.

Walther TC, Mann M. 2010. Mass spectrometry–based proteomics in cell biology. The Journal of

Cell Biology **190**, 491–500.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics **10**, 57–63.

Weigel D, Ahn JH, Blázquez MA, Ahn JH, Blázquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrándiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J. 2000. Activation Tagging in *Arabidopsis*. Plant Physiology **122**, 1003–1014.

Wu G, Otegui MS, Spalding EP. 2010. The ER-Localized TWD1 Immunophilin Is Necessary for Localization of Multidrug Resistance-Like Proteins Required for Polar Auxin Transport in *Arabidopsis* Roots. The Plant Cell **22**, 3295–3304.

Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, Wang X-C, Chen Q-J. 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biology **14**, 327

Yamamoto M, Wakatsuki T, Hada A, Ryo A. 2002. Gene expression profiling using improved SAGE. Rinsho byori. The Japanese journal of clinical pathology **50**, 52–60.

Yang F, Liang G, Liu D, Yu D. 2009. *Arabidopsis* MiR396 Mediates the Development of Leaves and Flowers in Transgenic Tobacco. Journal of Plant Biology **52**, 475–481.

Yang H, Liu X, Xin M, Du J, Hu Z, Peng H, Rossi V, Sun Q, Ni Z, Yao Y. 2016. Genome-Wide Mapping of Targets of Maize Histone Deacetylase HDA101 Reveals Its Function and Regulatory Mechanism during Seed Development. The Plant Cell **28**, 629–645.

Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabási AL, Tavernier J, Hill DE, Vidal M. 2008. High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. Science **322**, 104–110.

Yu S, Liao F, Wang F, Wen W, Li J, Mei H, Luo L. 2012. Identification of rice transcription factors associated with drought tolerance using the Ecotilling method. PLoS ONE **7**, e30765.

Zhang Y, Cai Y, Yu H, Li H. 2015. iTRAQ-Based Quantitative Proteomic Analysis Identified HSC71 as a Novel Serum Biomarker for Renal Cell Carcinoma. BioMed Research International **2015**, 802153–6

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome biology **9**, R137.

Zhou R, Benavente LM, Stepanova AN, Alonso JM. 2011. A recombineering-based gene tagging system for *Arabidopsis*. The Plant Journal **66**, 712–723.

LEAF GROWTH REGULATION AND THE ROLE OF ANGUSTIFOLIA3 IN *ARABIDOPSIS* AND *ZEA MAYS*

Nienke Besbrugge[1,2], Hilde Nelissen[1,2] and Geert De Jaeger[1,2]

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

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

Author contributions:

N.B. wrote the chapter; H.N and G.D.J. supervised and complemented the writing

ABSTRACT

Leaf growth is a complex biological process that is orchestrated by a myriad of proteins in both monocots and dicots. As leaves are the photosynthetic factories of the plant that provide food, feed, bio-materials and bio-energy to our society, unravelling the mechanisms that control leaf size and shape is a major field of study. In general, leaf growth can be seen as a subsequent series of four developmental phases (an initiation phase, a cell division phase, a transition phase and a cell expansion phase) in both monocots and dicots. The growth-regulating factor (GRF) family of transcription factors and the GRF-interacting factor (GIF) family of transcriptional co-activators have been identified in *Arabidopsis thaliana* to play a key role in cell division regulation during leaf growth. Recent analysis of the GRF-GIF duo in *Zea mays* further showed that the interaction between GRF and GIF1, also known as ANGUSTIFOLIA3 (AN3), is conserved between monocots and dicots, suggesting a similar regulatory role in leaf growth in both plant species. Here, we will review the different developmental processes that underlie leaf growth and provide an overview of the key components in these pathways, with a focus on the role of the GRF-AN3 duo during leaf growth in both *Arabidopsis* and *Zea mays*.

INTRODUCTION

Plant growth is a complex biological process orchestrated by a myriad of proteins and other biological molecules. Understanding how this process is regulated is a major objective in biology, as it is fundamental for agriculture and plant adaptation to the environment. Plant leaves are the centres of photosynthesis that produce biological energy through absorption of energy from sunlight. Besides providing a direct source of material for food, feed, bio-materials and bio-energy by the production of the building blocks for all biomass production, leaves also produce building blocks for the development of organs that consume these products, such as seeds and fruits. Leaves are therefore often referred to as source tissues, whereas the energy consuming organs are the sinks. The metabolism between these two tissues is stricktly coupled, and there is evidence that the growth capacity of the sink (and thus the agricultural yield) can be influenced by the photosynthetic capacity of the source. For example, increasing the leaf size of *Zea mays* through overexpression of PLASTOCHRON1 also resulted in a higher seed yield (Sun *et al.*, 2017). The photosynthetic capacity of the leaf is however greatly influenced by both size and shape of the final leaf, making leaf development an excessive field of study. In the past, most studies have focussed on leaf development in eudicots, using *Arabidopsis thaliana* as a model species, but more and more data are now also being generated in monocots, such as maize and rice. In general, leaf growth is seen as a series of four developmental phases that succeed one another: the initiation phase, a general cell division phase, a transition phase and a cell expansion phase. In *Arabidopsis,* an additional meristemoid division phase also influences the final leaf size (Gonzalez *et al.*, 2012). In maize, this phase does not exist, as monocots do not have meristemoid cells (Spangenberg, 2000). The

four general phases are tightly coordinated in both space and time to make sure that the final leaf size is achieved.

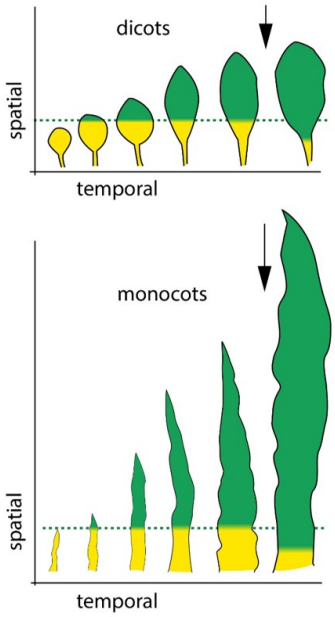


Figure 1. Schematic representation of leaf growth in monocots and dicots. Yellow parts represent actively dividing cells. The end of exponential growth is indicated with arrows. Figure taken from (Nelissen *et al.*, 2016)

Although monocots and dicots have distinct leaf shapes, the underlying mechanisms that drive leaf development appear to be partly conserved. In both plant types, leaf growth initiates at the leaf primordia, which flank the shoot apical meristem (SAM). At this time-point, all cells are actively dividing (*i.e.* division phase), creating groups of cells that will then migrate to their final destination in the leaf. After some time, cells at the tip of the leaf stop dividing and start to expand, whereas the cells at the base of the leaf keep proliferating (Fournier *et al.*, 2005; Gonzalez *et al.*, 2012). This transition of cell division to cell expansion creates a so-called cell cycle arrest front, which can be visualised in *Arabidopsis* by coupling the activity of the mitotic factor CYCLINB1;1 (CYCB1;1) promoter to CYCLINB1;1-GUS (Donnelly *et al.*, 1999). The arrest front moves gradually over time from the tip to the base of the leaf, providing a temporal characteristic on top of the spatial organisation to coordinate the underlying growth processes in *Arabidopsis* (Asl *et al.*, 2011; Andriankaja *et al.*, 2012, Beemster *et al.*, 2005). In monocot leaves, a similar linear organisation is present along the length axis of the leaf (dividing cells at the base of the leaf are followed by expanding cells and finally mature cells at the tip of the leaf) and the number of dividing cells also decreases over time (Fournier *et al.*, 2005; Avramova *et al.*, 2015; Sun *et al.*, 2017). This indicates that leaf growth is regulated in both plant types in a spatial and temporal manner, albeit to different extents (Figure 1). Detailed analysis of the influence of the cell division rate and the

duration of the cell division on leaf growth in *Arabidopsis* and maize further showed that these spatial and temporal pathways are regulated independently. For both species, a drop in cell division rate causes a decline in the overall growth rate of the leaf, suggesting that cell expansion alone is not sufficient to support maximum leaf growth in both monocots and dicots.

Besides the similar organisation of leaf growth, monocots and dicots also seem to utilise comparable molecular pathways in leaf development, as most of the genes that regulate growth occur in both plant types. For the purpose of this dissertation, the function of some of these genes will be discussed in more depth in the second part of this chapter, whereas the first part provides a broader view on the initial stages of leaf growth in general.

PART I - LEAF GROWTH

MERISTEM

In plants, all apical growth starts from a pluripotent stem cell population at the tip of the shoot, which is localised in the central zone (CZ) of the SAM (Figure 2). The actively dividing cells maintain their stem cell niche in the centre, whereas generating cells in the peripheral zones (PZ) that will differentiate into leaves and flowers. In eudicots, the SAM is composed of three layers: the epidermis (L1) and subepidermis (L2) which comprise the tunica, and the corpus or inner layer (L3). In order to maintain the identity of the pluripotent cells in the CZ, a tight regulation is provided by an interplay of the homeodomain TF WUSCHEL (WUS), expressed in the rib-zone (RZ), and the CLAVATA protein family (CLV 1, CLV2 and CLV3) expressed in the CZ of the SAM. CLV3, a small secreted ligand, moves from L1 and L2 to L3, and binds to receptor-like proteins CLV1 and/or CLV2, leading to inhibition of WUS activity.

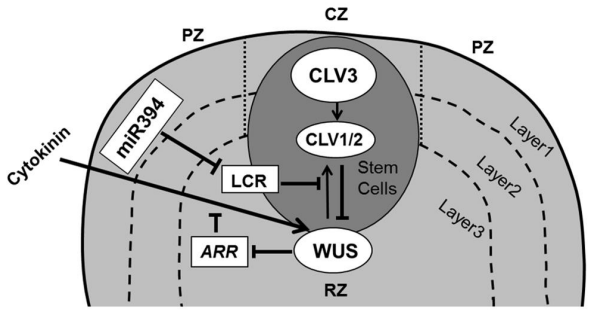


Figure 2. Stem cell maintenance in the shoot apical meristem (SAM). The SAM is arranged in three layers and three functional zones (central zone (CZ), peripheral zone (PZ) and rib zone (RZ)). The identity of the pluripotent cells in the meristem is maintained by the antagonistic action of WUS and CLV. Expression of WUS in the RZ leads to activation of CLV3. CLV3 then binds to CLV1/2, which inhibits WUS expression. (T shaped and pointed arrows indicate negative and positive regulation, respectively). Figure taken from Kalve et al., 2014.

In L3, WUS activity will lead to the formation of a non-cell autonomous signal that migrates to the stem cells where it induces CLV3 expression. Besides restricting its own accumulation, WUS relocalisation was also shown to function in the transcriptional repression of differentiation (Yadav *et al.*, 2011). The importance of this negative feedback loop is demonstrated by the conservation of the pathway in monocots, where orthologues of the CLV genes and WUS have been identified (Taguchi-Shiobara, 2001; Suzaki, 2004).

It has been proposed that changes in the size of the SAM could have an influence on the final leaf size, as a larger SAM can provide more founder cells for recruitment into the leaf primordia. Although a relationship between the SAM and leaf size through cytokinin (CK) signalisation was found to be present in *Arabidopsis* (Werner, 2003; Gordon *et al.*, 2009), there is no further evidence that supports this idea. In most cases, changes in SAM size will cause an elongation (when SAM is larger) or reduction in plastochron (*i.e.*, the duration of leaf production), but this is not a rule of thumb (Gonzalez *et al.*, 2012).

LEAF INITIATION

Although the spatial and temporal regulation of leaf growth is conserved between monocots and dicots, some differences are present that provide both clades their specific leaf identity. Leaves of dicotyledonous plants have a typical leaf base that is non-sheated, and a branched nervation pattern. Monocots, on the other hand, form leaves that contain veins in parallel organisation, and have a sheat at the leaf base that surrounds the stem at the node (Kaplan, 1973). In maize, the onset of leaf initiation is marked by the downregulation of the KNOTTED1 (KN1, in maize) TF. In *Arabidopsis*, a similar downregulation is seen for the orthologue of KN1, KNOTTED1-like homeobox protein (KNOX1). In both species, this downregulation is followed by recruitment of the founder cells that will form the incipient leaf primordium (P0) under the influence of auxin (Figure 3) (Sinha and Hake, 1992; Scanlon, 2003; Hay, 2006). A localised auxin maximum is created through influx via the AUXIN RESISTANT1 (AUX1) carrier and efflux by the PIN-FORMED1 (PIN1) transporter to the neighbouring cells with a higher auxin concentration (Guenot *et al.*, 2012). In *Arabidopsis*, the number of founder cells that is allocated to the primordium lies somewhere around 100, but varies among species. In *Zea mays*, on the other hand, about 40 meristematic cells are recruited to form the leaf founder initials (Li *et al.*, 2010). One could expect that a higher number of cells in P0 might cause an increase in the final leaf size due to the presence of more proliferating cells in the division phase. Indeed, two *Arabidopsis* mutants have been reported that support this idea. The first mutant, *struwwelpeter (swp)*, contains a truncated version of an RNA polymerase II transcriptional mediator that causes a reduction in both the amount of cells in the primordia and final leaf size (Autran *et al.*, 2002). This might indicate that the smaller leaf phenotype is due to a decreased number of founder cells. The *samba* mutant, on the other hand, has a larger SAM and enlarged leaves due to more initiator cells. Inactivation of SAMBA, which is a negative regulator of the Anaphase-Promoting Complex/Cyclosome (APC/C), leads

to stabilisation of A2-type cyclins during early development, favouring cell division (Eloy *et al.*, 2012).

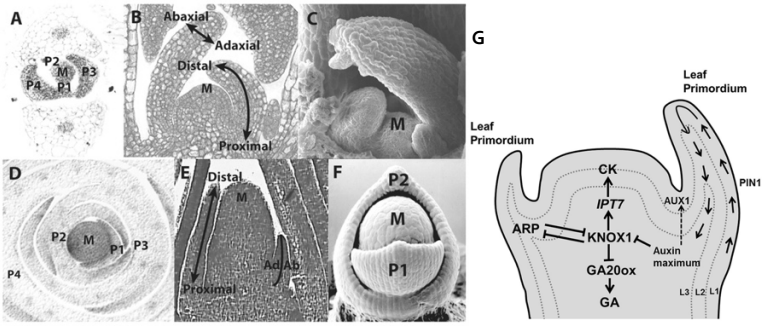


Figure 3. Development of leaf primordia from the shoot apical meristem (SAM). (A) and (B) Transverse and Longitudinal sections of the SAM of *Arabidopsis,* **(D) and (E)** Transverse and Longitudinal sections of the SAM of *Zea mays*. Order of leaf primordia is indicated, with P1 being the youngest, and increasing age with P2 and onward. Proximal-distal and Abaxial-Adaxial (Ab/Ad) axes are indicated. Note the spiral-like phyllotaxis of leaves in *Arabidopsis*, whereas *Zea mays* leaves encircle the SAM and have an alternating phyllotaxis. **(C) and (F)** Scanning electron micrographs (SEM) of the SAM of respectively *Arabidopsis* and *Zea mays* showing the difference in spatial organisation of primordia compared to each other and to the SAM. M: Shoot apical meristem. Figure taken from (Chitwood *et al.*, 2007). **(G)** Schematic representation of leaf initiation in *Arabidopsis*. Auxin accumulates via PIN1/AUX1-mediated polar auxin transport in the flanks of the SAM, triggering primordia development. Emergence of the primordium is regulated via ARP proteins. In the meantime, the key regulator KNOX1 maintains stem cell identity and positively regulates cytokinin through *IPT7.* It also negatively influences GA signalling via GA20 oxidase. The different layers of the SAM are indicated L1-L3. Figure taken from (Kalve *et al.*, 2014).

In *Zea mays*, other regulators of the meristem size, such as the KN1 counteracting bZIP TF FASCIATED EAR4 (FEA4)(Pautler *et al.*, 2015) and the MALE STERILE CONVERTED ANTHER1 (MSCA1) glutaredoxin protein (Yang *et al.*, 2015), have also been found, although it is not clear if they also influence the leaf size. FEA4 is thought to be involved in the acceleration of cell differentiation in the peripheral zone through inhibition of the meristem-promoting factors (Pautler *et al.*, 2015). Interestingly, a novel pathway containing MSCA1 and FEA4 for SAM regulation was proposed, based on their interaction in a Y2H screen (Yang *et al.*, 2015). The stem cells in the SAM and the cells in the leaf primordium are both dividing, but have a distinct identity. Although not much is known about the underlying regulatory networks, analysis of leaf growth in *Arabidopsis* showed the importance of the interplay between the KNOX1 and the ASYMMETRIC LEAVES1 (AS1)/ROUGH SHEAT2 (RS2)/PHANTASTICA (PHAN) (ARP) TF families (Figure 3). *KNOX*, which is normally expressed throughout the meristem, is absent at the site of organ initiation (Kerstetter *et al.*, 1994; Long *et al.*, 1996; Scofield and Murray, 2006). On the other hand, *AS1* mRNA, which is not expressed in the meristem, is transcribed in the primordia-forming cells (Martienssen *et al.*, 2000). ARP proteins are thus the drivers of leaf primordium differentiation. In regard to the hormonal signals involved, CK biosynthesis is increased (Yanai *et al.*, 2005) by KNOX1, the TF that is necessary for the maintenance of undifferentiated cells in the SAM (Scofield

and Murray, 2006). When auxin levels are high, the activity of KNOX1 decreases, leading to a reduction in CK biosynthesis (Su *et al.*, 2011) through inactivation of *IPT7* (an isopentenyl transferase enzyme that synthesises CK). Other KNOX genes that are necessary for the initiation of leaf growth are *BREVIPEDICELLUS (BP)*, which is down-regulated by a combination of high auxin levels and AS1 expression (Hay, 2006), and *SHOOTMERISTEMLESS* (STM), an antagonist for the ARP proteins in the regulation of leaf primordia production. This is demonstrated by the fact that *as1* and *as2* mutants of *Arabidopsis* have smaller and rounder leaves (Bowman and Eshed, 2000), whereas plants that lack a functional STM are inhibited in the production of both SAM and leaves (Long *et al.*, 1996). A third hormone involved in the maintenance of stem cell identity in the SAM is gibberellin (GA). To stop the cells from differentiating, GA levels have to be low, which can be achieved by downregulation of the GA20 oxidases rate-limiting enzymes involved in GA production (Sakamoto, 2001; Jasinski *et al.*, 2005; Veit, 2008; Gordon *et al.*, 2009).

LEAF GROWTH

After the initiation phase, leaf growth in monocots and dicots continues through a phase of general cell division, followed by a transition and an expansion phase. During the cell division phase, cells are actively proliferating (one assumes for a constant period in time). This links cell division activity to final leaf size, indicating that faster progression through the cell cycle will influence the total number of cells in the leaf. Strikingly, most proteins involved in the cell cycle are conserved among eukaryotes, and their interactions are similar in yeast, humans and plants. Complexes of cyclins (CYC) and cyclin-dependent kinases (CDK), respectively the regulatory and the catalytic subunits, are the main regulators of cell cycle progression. CDKs are activated throughout the cell cycle by the CDK-activating kinases CDKD and CDKF, whereas their function is inhibited by Kip related proteins (KRPs) (Figure 4). Two crucial check points of cell cycle progression are the transition from the G1 to S phase in the beginning, and the exit of mitosis at the end of the cycle. For G1/S, activated complexes consisting of CDKA and CYCD will phosphorylate the E2F/DP (dimerisation protein) heterodimeric complex. E2F/DP then dissociates from its negative regulator, RBR (retinoblastoma-related protein). The activated E2F/DP complex relocalizes and stimulates the transcription of genes involved in DNA duplication. Once the DNA is duplicated, the activity of CDKA and CDKB, which is controlled by the WEE1 kinase, together with CYCA, CYCB and CYCD will lead to initiation of the G2 phase (Kalve *et al.*, 2014). During this G2 phase, cells prepare for mitotic division (M phase). When mitosis is completed, the proteolytic activity of the APC/C complex, an E3 ubiquitin ligase complex, will lead to degradation of the cyclins and exit from mitosis. Overexpression of subunits of this complex, such as APC10 and CDC27a (cell division cycle protein 27 homologue a) increases the cell division rate in *Arabidopsis* and tobacco (*Nicotiana tabacum*) respectively, leading to an enhanced final leaf size (Eloy *et al.*, 2011). Important regulators of the APC/C complex are the CCS52 (CELL CYCLE SWITCH52) and CDC20 activators, and the inhibitory protein ULTRAVIOLET-B-INSENSITIVE4 (UVI4) (Bao and Hua, 2014).

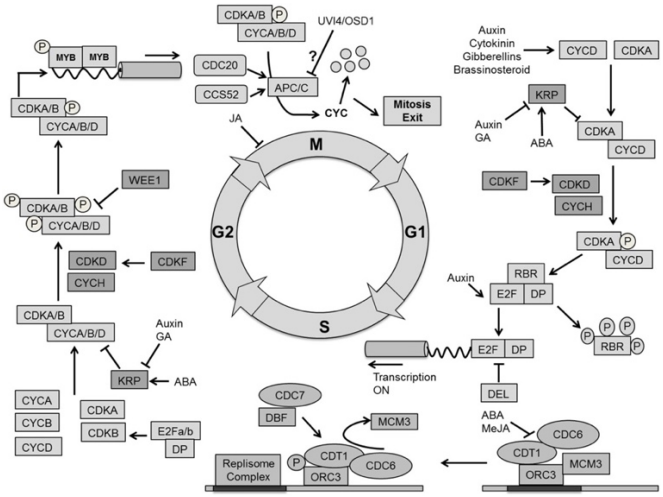


Figure 4. Cell cycle regulation in plants. Progression through the four phases of the cycle (G1, S, G2, M) is regulated by cyclin-dependent kinases (CDKs), which are successively activated via binding with cyclins and phosphorylation by CDK-activating kinases (CDKD and CDKF), and deactivated through KRPs. Phosphorylation of RBR proteins by CDKA-CYCD triggers the release of TF E2F, which will activate S phase related genes and aids G1-S phase transition. CDKA/B and CYCA/B/D, on the other hand, regulate the transition of G2-M. Here, phosphorylation of the CDK complex by WEE1 causes inactivation. To exit mitosis, CYCs are degraded by the Anaphase-Promoting Complex/Cyclosome (APC/C) that is activated through CCS52 and CDC20. The positive/negative regulatory role of plant phytohormones (auxin, cytokinin, gibberellins (GA), brassinosteroids, abscisic acid (ABA) and methyl jasmonate (MeJA)) in cell cycle progression is indicated at several points. Figure taken from (Kalve *et al.*, 2014).

Another E3 ubiquitin ligase complex, the SCF (Skpl–cullin–F-box-protein) also regulates the cell cycle. It targets the inhibitory KRPs (Hershko, 2005; Ren *et al.*, 2008) as well as the transcriptional repressor complex E2Fc/DP/RBR (Inzé and De Veylder, 2006). Association of the SCF with another F-box protein, SLEEPY1 (SLY1), leads to the degradation of DELLA proteins. DELLAs are nucleair repressors of growth in the absense of GA (Dill *et al.*, 2001). Binding of SLY1 to the DELLA proteins upon GA signalling leads to proteolytic degradation of the latter, and loss-of-function *sly1* mutants accumulate high levels of DELLA proteins as they are GA insensitive (Dill, 2004). They have a dwarf phenotype due to a decreased cell proliferation rate (Achard *et al.*, 2009). On the other hand, DELLA proteins also influence leaf growth by negatively controlling the expansion rate of differentiated cells. As such, destabilisation of DELLA proteins by overexpressing the Gibberellin 20 oxidase1 (GA20OX1) rate-limiting enzyme in GA production will cause an increase in both cell number and cell size. This leads to the formation of larger leaves in both monocots and dicots (Gonzalez *et al.*, 2010).

Other players that influence the final leaf size through regulation of cell cycle progression are members of the GRF INTERACTING FACTOR (GIF) family of transcriptional co-activators (discussed in part II of this chapter) (Lee *et al.*, 2009).GIFs are positive regulators of CYCB1;1 expression, and *Arabidopsis* plants mutated in all three GIF proteins, *i.e.*, *gif1/gif2/gif3* triple mutants, have a reduced

expression of CYCB1;1 and other cell cycle related genes (Lee *et al.*, 2009). As a result, both the cell proliferation rate as the duration of the cell division phase are decreased, and the final leaf area is smaller (Lee *et al.*, 2009).

TRANSITION FROM CELL DIVISION TO CELL EXPANSION IN A SPATIO-TEMPORAL MANNER

Final leaf size is tightly linked to the number of cells generated during the division phase. Therefore, besides the rate of cell division, the duration of the cell proliferation phase will also have a great influence on the organ size. In *Arabidopsis,* the transition from division to expansion (and thus cell cycle exit) is generated via a cell cycle arrest front, which moves in a spatio-temporal manner from the tip to the base of the leaf (Andriankaja *et al.*, 2012). The process, regulated by a long list of factors, has been extensively studied, but to keep an overview only the key regulators are discussed below (Figure 5) (see (Kalve *et al.*, 2014) for an extended review).

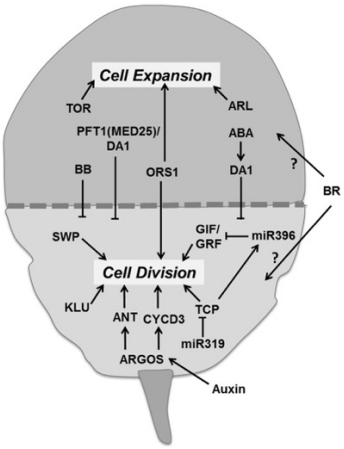


Figure 5. Regulatory mechanisms during transition of cell proliferation to cell expansion in *Arabidopsis*. The dashed line indicates the transition between both growth processes, thereby separating the regulatory proteins. Auxin-induced ARGOS activates TF ANT and $CYCD_3$, which promotes cell division. Other positive regulators of cell proliferation are miR396 target GIF/GRF, miR319 target TCP, SWP and KLU. Promoters of cell expansion are the TOR pathway as well as ARL. On the other hand, BB, MED25 and DA1 all play a regulatory role in the duration of proliferation. Factors that influence both growth processes also exist, such as ORS1 that is shown here. Some phytohormones also regulate transition between both phases to some extent, as is demonstrated here for ABA and DA1. Unknown mechanisms are indicated with arrowheads. Figure taken from (Kalve *et al.*, 2014)

The first factor that is involved in the transition from proliferation to expansion is AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS). Members of the ARGOS gene family (respectively four and eight in *Arabidopsis* and *Zea* mays) are integral membrane proteins that localise to the ER and Golgi (Rai *et al.*, 2015; Shi *et al.*, 2015). *Arabidopsis* plants overexpressing ARGOS genes have larger leaves with more cells, and are less sensitive to ethylene (Shi *et al.*, 2015). The ORGAN

SIZE RELATED1 (ORS1) protein of *Arabidopsis*, which contains a domain that is conserved in ARGOS and ARGOS-LIKE (ARL), was also shown to have a positive effect on cell proliferation and expansion during leaf growth (Hu, 2003; Feng *et al.*, 2011). In maize, the increase in cell number due to overexpression of *ARGOS1* (*ZAR1*) was not only seen in the leaf, but also in the stalk, the ear and the grain yield. Besides affecting growth, high levels of ZAR1 also promote drought stress tolerance (Guo *et al.*, 2014), and ectopic expression of ZAR1 in *Arabidopsis* was shown to influence the perception or signalisation of ethylene (Shi *et al.*, 2015) by interacting directly with the ethylene receptor signaling complex (Shi *et al.*, 2016). Overexpression of ARGOS8, another member of the maize ARGOS family, also provided this drought stress tolerance, together with an increase in grain yield (Shi *et al.*, 2015). Another group of proteins that play a crucial role in leaf growth in monocots and dicots is the TCP (TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1 and 2) family of TFs, named after its first four characterised members in respectively maize (*Zea mays*), snapdragon (*Antirrhinum majus*) and rice (*Oryza sativa*). TCP activity regulates the expression of two miRNAs: miR396, which targets members of the GROWTH REGULATING FACTOR (GRF) TF family (see below), and miR319, which helps to control the accumulation level of TCP family members (Palatnik *et al.*, 2003; Rodriguez *et al.*, 2009). Interestingly, an increase in TCP4 expression in *Arabidopsis*, which can be counteracted by upregulation of miR319, leads to a decrease in the final leaf size (Rodriguez *et al.*, 2009) by inducing miR396 mediated inhibition of GRF expression. Other members of the family, such as the CINCINNATA TCPs, are also known to control cell cycle arrest in *Arabidopsis* (Nath, 2003; Palatnik *et al.*, 2003; Ori *et al.*, 2007).

The duration of proliferation is also regulated by the putative ubiquitin-binding protein DA1 (Li *et al.*, 2008) and its enhancer BIG BROTER (BB) (also known as ENHANCER OF DA1-1 (EOD1)), an E3 ubiquitin ligase (Disch *et al.*, 2006) (Figure 5). In addition to negatively controlling the moment of transition from proliferation to expansion in a synergistic manner, DA1 and BB were found to promote leaf senescence in *Arabidopsis* (Vanhaeren *et al.*, 2017). Moreover, maize plants overexpressing a mutated DA1 produced larger seeds, suggesting that the function of the protein is conserved among species (Wang, 2012). In monocots, GRAIN WIDTH AND WEIGHT2 (GW2) was also shown to negatively affect the grain size (Song *et al.*, 2007; Su *et al.*, 2010). GW2 is an orthologue of another ubiquitin E3 ligase protein (DA2) that works in synergy with DA1 in *Arabidopsis* (Xia *et al.*, 2013). Two members of the CYP78A cytochrome P450 family, namely KLUH (KLU)/CYP78A5 and PLASTOCHRON1 (PLA1), on the other hand, are positive regulators of leaf growth. Conserved across species, KLU is thought to generate a mobile growth signal that promotes leaf growth and differs from the standard phytohormones. Plants containing a defective KLUH gene produce smaller leaves due to a shorter proliferation phase (Stransfeld *et al.*, 2014). In rice and maize, PLA1 was reported to have a similar effect on vegetative growth and leaf initiation (*i.e.* plastochron) as KLUH, and its mode of action in maize was shown to be GA-independent (Sun *et al.*, 2017).

Hormones that play a role in the regulation of leaf growth are auxin, GA and brassinosteroids (BRs), but the underlying mechanisms are still poorly understood (Figure 5). *Arabidopsis* plants overexpressing the BR receptor BRI1 (BRASSINOSTEROID INSENSITIVE1) were reported to have larger leaves (Gonzalez *et al.*, 2010; Oh *et al.*, 2011). Conversely, a growth reducing effect was seen in maize plants in which the *BRI1* expression was down-regulated (Kir *et al.*, 2015). Also in maize, a peak in GA accumulation was shown to mark the position of the transition zone in the leaf (Nelissen *et al.*, 2012), and ectopic expression of the GA biosynthesis rate-limiting enzyme GA20OX1 was shown to generate larger leaves in both *Arabidopsis* and *Zea mays* (Gonzalez *et al.*, 2010; Nelissen *et al.*, 2012).

ESTABLISHING THE ADAXIAL/ABAXIAL LEAF FATE

Upon lamina outgrowth, an asymmetric identity is created that provides a polarity to the developing leaf. The side of the leaf that faces the SAM produces the adaxial part, whereas the cells facing away from the SAM gain the abaxial fate. The underlying mechanisms that are required to establish this dorsoventral polarity are still not fully understood, but some of the antagonistic TFs involved in the regulation of abaxial/adaxial cell fate have been identified.

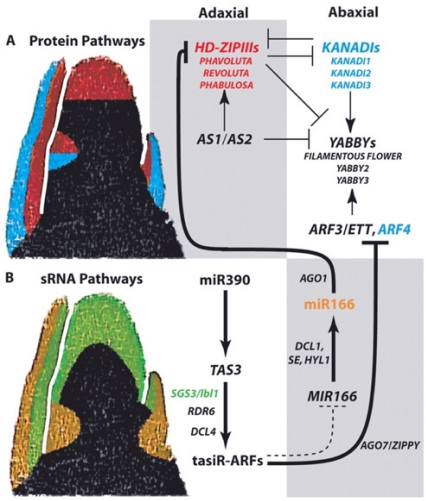


Figure 6. Molecular mechanisms that control adaxial/abaxial patterning in *Zea mays* leaves. Protein regulated pathways **(A)** versus small RNA (sRNA)-regulated pathways **(B)** are subdivided into molecular drivers that confer adaxial or abaxial cell fates. Putative regulatory interactions are indicated by dashed lines. Colours represent the meristematic expression pattern of the factor with the corresponding colour. Figure taken from (Chitwood *et al.*, 2007)

The first mutant characterised in *Arabidopsis* was *phabulosa-1* (*phb-1d*). Leaves of this mutant were found to be radially symmetrical instead of flat, and failed to produce a leaf blade (Thoma and Chandler, 2015). PHB belongs to the class III homeodomain-leucine zipper (HB-ZIPIII) family of TFs and determines the adaxial

identity together with its family members PHAVULOTA (PHV) and REVOLUTA (REV) (Figure 6) (McConnell *et al.*, 2001).

On the abaxial side, expression of the KANADI(KAN) TF, a Golden2/*Arabidopsis* response-regulator/O starvation/acclimatisation response (Psr1;GARP) regulator (Kerstetter *et al.*, 2001; Eshed *et al.*, 2001), and the YABBY (YAB)/FILAMENTOUS FLOWER (FIL) gene family (Eshed, 2004) are the main drivers of identity. To maintain the abaxial/adaxial differentiation, a feedback loop has to be governed between these two classes of genes: KAN and YABBY (class I) will repress the expression of the HD-ZIPIII abaxial determinators in the abaxial part, whereas the HD-ZIPIII proteins (class II) inhibit the expression of KAN genes in the adaxial side of the leaf (Tsukaya, 2013*a*; Kalve *et al.*, 2014). Besides their function in the adaxial patterning of the leaf primordia, REV and PHB are also expressed in the tip of the SAM. This suggest that they also play a role in the meristem, and might even regulate the coordination between both processes. This connection is partly confirmed by the fact that plants carrying a dominant mutation in the HB-ZIPIII genes have a larger SAM, which can be counteracted by mutations in the STM gene (Thoma and Chandler, 2015).

Other important regulators of the adaxial leaf fate are ASYMMETRIC LEAVES1 (AS1) and AS2. AS1 was mentioned previously for its role in the inhibition of the KNOX1 expression during leaf primordium initiation. In *Arabidopsis*, mutations in AS1 or AS2 do not demolish adaxial leaf identity, but result in plants with downwarly curled leaves that sometimes contain leaflet like structures (Tsukaya and Uchimiya, 1997; Martienssen *et al.*, 2000). However, the AS1/AS2 complex has been shown to regulate a variety of other proteins involved in adaxial/abaxial patterning, such as, for example, GIF1 (Horiguchi *et al.*, 2011; see below). It was also shown that the complex binds through AS1 to the promoter of abaxially expressed AUXIN RESPONSE FACTOR3 (ARF3)/ETTIN (ETT), causing its repression. A similar genetic interaction is seen between AS1/AS2 and YAB5, KNAT1, KNAT2, and KAN, although direct binding of their respective promoter sequences has not been reported.

Three small RNAs, miR165/166, miR390 and transacting small interfering RNA (tasiRNA) tasiR-ARF also help to regulate the establishment of leaf patterning, but function upstream of the factors discussed above (Juarez, 2004; Nogueira *et al.*, 2007). Binding of miR165/166 to ARGONAUTE1 (AGO1) at the adaxial side stimulates the cleavage of HD-ZIPIII transcripts. Abaxialy expressed miR390 on the other hand will bind to AGO7, which leads to the processing of TAS3 to tasiR-ARF, a negative regulator at the post transcriptional level of ARF3 and ARF4 expression. This induced degradation is partly counteracted by a positive feedback loop created by YAB3 controlled KAN1 and ARF4 expression (Bonaccorso *et al.*, 2012). Interestingly, binding of miR390 to AGO7 recruits other components of the RISC (RNA-induced silencing complex): DICER-LIKE4 (DCL4), SUPPRESSOR-OF-GENE SILENCING3 (SGS3) and RNA-DEPENDENT RNA POLYMERASE6 (RDR6)(Peragine, 2004; Allen *et al.*, 2005; Xie *et al.*, 2005). In *Arabidopsis*, four *TAS* gene families (TAS1-4) have been reported, whereas only one family (tas3) is found

in maize (Nogueira *et al.*, 2007). Post-transcriptionally processed TAS3 is further cleaved into its 21-bp tasiRNAs by a complex consisting of SGS3-RDR6-DCL4. Orthologues of SGS3 and AGO7 in maize, are LEAFBLADELESS1 (LBL1) and RAGGED SEEDLING2 (RGD2), respectively. However, whereas adaxial expression of tasiR-ARF is necessary to inhibit adaxial expression of *ARF3* in *Arabidopsis*, *rgd2* LOF mutants in maize did not accumulate *ARF3* mRNA in the adaxial side of the leaf. This suggest that other signals are involved in maintaining the dorsoventral identity of the leaf in maize (Douglas *et al.*, 2010).

Besides the diversification of the adaxial/abaxial sides, cells in the middle of the leaf that form the boundaries of both surfaces of the leaf have a specific cell fate. However, knowledge on the underlying mechanism is still lacking. At present only two factors have been reported to be involved, namely WUS-RELATED HOMEOBOX (WOX) 1 and 3, and their expression is inhibited by KAN (Nakata *et al.*, 2012; Nakata and Okada, 2014).

CELL EXPANSION

As soon as cells exit their proliferation phase, leaf growth is further driven by cell wall loosening, together with *de novo* synthesis of cell wall components and turgor-based cell expansion (Cosgrove, 2005). Although similarly to the division phase duration and rate of the cell expansion process could influence the final leaf size, only a hand full of cell expansion controlling factors have been reported. For example, upregulation of expression of EXPANSIN10 (EXP10) in *Arabidopsis*, a cell wall-associated protein belonging to the EXPANSIN protein family (Sampedro and Cosgrove, 2005), increases its cell wall loosening activity, leading to larger cells and an enlarged final leaf size (Cho and Cosgrove, 2000). In maize, repression of the *ZmEXPB6* expression leads to a reduction in leaf growth under salt stress conditions (Geilfus *et al.*, 2015).

Another driver of cell expansion is endoreduplication, an alternative version of the cell cycle in which cells continue to replicate their genome without undergoing subsequent mitosis (Breuer *et al.*, 2010). Conserved among plant species, endoreduplication links the size of the cell to its ploidy level (Melaragno, 1993; Kondorosi *et al.*, 2000; Sugimoto-Shirasu and Roberts, 2003; Tsukaya, 2013*b*), but it is not necessary for cell expansion (Veylder *et al.*, 2001).

Other factors for which a positive role in cell expansion has been suggested in *Arabidopsis* are ARGOS-LIKE (ARL), the TARGET OF RAPAMYCIN (TOR) kinase, and the ZINC FINGER HOMEODOMAIN5 (ZHD5) TF, although their mode of action in cell enlargement remains unknown (Hu *et al.*, 2006; Deprost *et al.*, 2007; Hong *et al.*, 2011).

MERISTEMOID DIVISION

In *Arabidopsis*, leaf growth is finalised by a meristemoid division phase. Although most epidermal cells start to differentiate at a certain point during the cell expansion phase, meristemoid cells maintain their active proliferation status for

several additional rounds. These extra divisions are asymmetrical and generate cells that will differentiate into stomatal guard cells or provascular cells (Nadeau, 2002; Fisher and Turner, 2007; Peterson *et al.*, 2010). The importance of the contribution of meristemoid division to the final leaf size is exemplified by the production of the stomatal lineage, which not only provides the stomata but also generates a substantial number of epidermal cells known as pavement cells (three per meristemoid cell) (Geisler *et al.*, 2000; Bergmann and Sack, 2007). Downregulation of the PEAPOD (PPD) genes, which are DNA-binding proteins that repress meristemoid proliferation by regulating the occurrence of a second cell cycle arrest front, results in larger, dome-shaped leaves (White, 2006). Hence, PPD is thought to control the expression of D-type cyclins, which are important factors needed for cell cycle progression (Gonzalez *et al.*, 2015). Interestingly, there are no orthologues of PPD in monocots, coinciding with the absence of meristemoid cells in the latter. Rather, stomata are formed by cell division of an epidermal cell via a regular pattern, which results in a large cell towards the base of the leaf and a small cell towards the leaf tip. As the smaller cell (also known as the initial cell) divides into two symmetrical guard cells, an alternating pattern of stomatal initials and neighbouring cells is produced within the linear cell files (Spangenberg, 2000).

PART II: THE ROLE OF THE GIF AND GRF PROTEIN FAMILIES IN LEAF GROWTH

Besides the many regulators discussed above that are involved in leaf growth, two protein families have been extensively studied for their role in plant development: the GROWTH REGULATING FACTOR (GRF) TF family and the GRF INTERACTING FACTOR (GIF) family of transcriptional co-activators. Both groups are essential for proper plant growth, and are expressed in several parts of the plant during development, such as the leaves (van der Knaap *et al.*, 2000; Kim *et al.*, 2003; Horiguchi *et al.*, 2005; Kim and Lee, 2006) or the roots. They were shown to interact with each other in complexes, and are both controlled indirectly (GIFs) or directly (GRFs) at the transcriptional level by miR396.

The first GRF that was identified is the rice OsGRF1, a plant-specific TF shown to be upregulated upon GA induction (van der Knaap *et al.*, 2000). Since then, more members of the GRF family have been found not only in rice, but also in *Arabidopsis*, *Zea mays*, *Glycine max*, *Brassica napus*, *Solanum tuberosum*, *Physcomitrella patens* and all other land plants for which the genomes have been sequenced (Zhang *et al.*, 2008; Yang *et al.*, 2009; Osnato *et al.*, 2010; Liu *et al.*, 2012; Kim *et al.*, 2012; FİLİZ *et al.*, 2014*a*,*b*; Wang *et al.*, 2014; Kuijt *et al.*, 2014). Their interaction with the GIF family is also conserved among plant species (Kim and Kende, 2004; Lee *et al.*, 2009; Lee and Kim, 2014; Nelissen *et al.*, 2015). For the purpose of this thesis, the function of both protein families and their interaction patterns during leaf development are described in depth for *Zea mays* and *Arabidopsis thaliana.*

PROTEIN STRUCTURE AND MOLECULAR FUNCTIONS OF GRFS AND GIFS.

GRF

At present, nine and 17 members of the GRF family have been identified in *Arabidopsis* and *Zea mays*, respectively. Their size ranges from 20 kDa to 60 kDa, which mostly depends on the length of the C terminus. The N-terminal region of GRF proteins typically harbours two conserved parts: a QLQ (glutamine, leucine, glutamine, IPR014978) domain and a WRC (tryptophan, arginine, cysteine, IPR014977) amino acid domain (van der Knaap *et al.*, 2000; Choi, 2004; Zhang *et al.*, 2008)(see Figure 7). The QLQ domain consists of approximately 36 amino acids of which some (besides the QX_3LX_2Q residues) are conserved and aromatic/hydrophobic or acidic in nature. The domain shares great similarity with the QLQ domain in the N-terminal part of SWITCH2/SUCROSE NONFERMENTING2 (SWI/SNF), a chromatin remodelling ATPase found in *Saccharomyces cerevisiae* that uses the QLQ domain as an interface for interaction with other subunits of the chromatin-remodelling complex (Treich *et al.*, 1995). It was therefore thought that the high conservation of the QLQ domain is important for the formation of PPIs in plants (van der Knaap *et al.*, 2000; Kim *et al.*, 2003; Hoe Kim and Tsukaya, 2015). Indeed, experimental analysis of the interaction of AtGRF and AtGIF in *Arabidopsis* showed that they bind through the QLQ domain (Kim and Kende, 2004; Horiguchi *et al.*, 2005). The WRC domain, on the other hand, contains about 44 amino acids that are clustered into two groups: a C_3H motif, made up of three cysteines and a single histidine residue ($CX_9CX_{10}CX_2H$), and a stretch of mostly basic amino acids that shows little variability among the different GRFs. Given that the C_3H motif is similar to the one found in the barley HRT transcriptional repressor that binds to the GA response element (Raventós *et al.*, 1998; van der Knaap *et al.*, 2000), it has been suggested that this C_3H motif functions as a DBD. This was confirmed for the GRFs through a yeast one-hybrid (Y1H) assay and electrophoresis mobility shift essay (EMSA) in rice, barley and *Arabidopsis* (Osnato *et al.*, 2010; Kim *et al.*, 2012; Kuijt *et al.*, 2014). The basic amino acids, on the other hand, might aid in targeting the proteins to the nucleus, as most GRF proteins were found to have a nuclear localisation (van der Knaap *et al.*, 2000; Osnato *et al.*, 2010; Kim *et al.*, 2012; Kuijt *et al.*, 2014; Liu *et al.*, 2014). Interestingly, all nine members of the *Arabidopsis* GRF family contain both domains at their N terminus (Kim *et al.*, 2003), whereas in maize only 14 out of 17 ZmGRFs were reported to have a QLQ and/or WRC domain (Omidbakhshfard *et al.*, 2015).

The C-terminal part of GRF proteins is more variable, but transactivation assays using the Y2H method showed that the C-terminal domain acts as transactivating domain in both *Arabidopsis* and rice (Choi, 2004; Kim and Kende, 2004; Liu *et al.*, 2014). Moreover, no transactivation activity was seen for ZmGRF10 from *Zea mays*, which has a very short C-terminal domain (Wu *et al.*, 2014). Strikingly, many AtGRFs were found to be functionally redundant in the regulation of plant development and growth, yet still they are able to perform unique molecular and biological functions. It is not yet clear how this molecular specificity is established, as the N-terminal QLQ and WRC domains are so highly conserved, but it is suggested that the C terminus (partly) confers this function.

All the above clearly indicates that GRFs are TFs that contain a plant-specific DBD, but whether they act as positive or negative regulators of transcription is still being investigated. It was long thought that GRFs were transcriptional activators. However, recent studies provide a different view (Kim *et al.*, 2012; Kuijt *et al.*, 2014). A LOF mutation in AtGRF7, for example, which binds the promoter sequence of Dehydration-responsive element binding protein2A (DREB2A), resulted in a significant increase in the transcript levels of its target genes (Kim *et al.*, 2012). Another possibility is that the function of GRFs can vary in different genomic contexts, suggesting also a regulatory role for its interaction with for example members of the GIF family (Hoe Kim and Tsukaya, 2015).

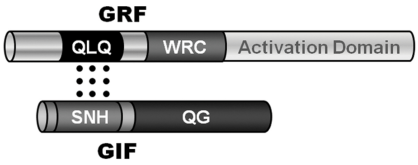


Figure 7. Schematic overview of GRF and GIF protein features. Dotted lines represent interactions between both domains. Figure taken from (Hoe Kim and Tsukaya, 2015)

GIF

In *Arabidopsis* and *Zea mays*, the GIF family contains three members: GIF1/ANGUSTIFOLIA3 (AN3), GIF2 and GIF3. GIF proteins are homologues of the transcriptional coactivator synovial sarcoma translocation protein (SYT), in humans (Thaete *et al.*, 1999), which lacks a DBD and interacts with human SWI/SNF ATPases, such as BRAHMA (BRM) and BRM-RELATED GENE1 (BRG1) (Brett, 1997; Santos, 1997; Kato *et al.*, 2002; Aizawa, 2004). They are involved in *Arabidopsis* in the regulation of cell proliferation during leaf development as well as in the maintenance of pluripotent meristematic cells in reproductive organs (Kim and Kende, 2004; Lee *et al.*, 2009; Lee and Kim, 2014). The first two members of the family in *Arabidopsis*, AtAN3 and AtGIF2, were identified as interactors of AtGRF1 via a Y2H screen, and their interaction was found to be mediated by the QLQ domain (Kim and Kende, 2004). Since then, interaction of AtAN3 with other GRFs, such as AtGRF3 and AtGRF5, has been repeatedly reported (Horiguchi *et al.*, 2005; Vercruyssen *et al.*, 2014; Debernardi *et al.*, 2014). Interestingly, the interaction between the two protein families was also confirmed in rice and maize, indicating that their molecular functions might be conserved between monocots and dicots (Liu *et al.*, 2014; Wu *et al.*, 2014; Nelissen *et al.*, 2015)(see below).

The GIF proteins, which have a molecular mass of ~20kDa, contain a conserved stretch of 54 amino acid residues located at the N terminus. Similar to the SNH domain found in SYT (Kim and Kende, 2004; Horiguchi *et al.*, 2005), it provides the interface for the interaction with the QLQ domain of GRFs (Figure 7). The C-terminal part of GIFs differs a bit more from their SYT orthologue, but is equally enriched for glutamine (Q) and glycine (G) and therefore comparable with the QPGY domain of SYT. Although an obvious nuclear localisation signal is lacking

from its protein sequence, AtAN3 was found to be preferentially localised in the nucleus, where it can be seen as speckles (Kim and Kende, 2004). Moreover, the nuclear localisation of complexes formed by certain GRF and GIF pairs has been reported in both *Arabidopsis* and rice (Liang *et al.*, 2014; Liu *et al.*, 2014). Additionally, AN3 works as a transactivator, similar to SYT in humans, which was shown through Y2H and transient expression studies (Kim and Kende, 2004).

Post transcriptional control of GRF expression by miR396

The accumulation of transcripts of most *GRF* genes has been found to be regulated by miR396-mediated silencing. In *Arabidopsis*, the precursor gene of miR396, *MIR396*, encodes two miRNAs (miR396a and miR396b) that target and cleave the AtGRF mRNA of all AtGRFs except AtGRF5 and AtGRF6 (Jones-Rhoades *et al.*, 2006; Liu *et al.*, 2009). Plants overexpressing MIR396 were found to be smaller, which is consistent with the reported growth enhancing function of the GRF family (Rodriguez *et al.*, 2009). Interestingly, decreased levels of AtGRF mRNA were also reported to have a reducing effect on AtAN3 transcription (Rodriguez *et al.*, 2009), an effect that was also seen in rice (Liu *et al.*, 2014). In maize, all GRF members, with the exception of ZmGRF4 and ZmGRF10, contain a miR396 target site (Nelissen *et al.*, 2015). The function of miR396 mediated post transcriptional regulation during leaf growth is discussed below.

HOW DO GIFS REGULATE LEAF GROWTH?

The growth-regulatory effect of GIFs was first discovered in the *Arabidopsis an3* LOF mutant. Whereas *gif2* and *gif3* single-mutants did not show any phenotypical alterations, *an3* plants displayed smaller and narrower leaves, containing fewer cells than their wild-type counterparts (Kim and Kende, 2004; Horiguchi *et al.*, 2005). The synergistic effect of AN3, GIF2 and GIF3, as seen by the even stronger small plant phenotype of the *Atgif1/2/3* triple mutant, also indicated that all three GIFs are involved in positive regulation of growth, and that their roles are (partly) redundant (Lee *et al.*, 2009). To investigate whether the AtGIFs influence the cell division rate or the duration of the cell division phase, leaf primordia of the triple mutant were subjected to kinematic analysis. Based on the low cell number due to a decreased cell division rate and the observation of a precocious exit from the cell division phase during leaf growth, it was concluded that AtGIFs control both processes (Ferjani *et al.*, 2007; Lee *et al.*, 2009). Congruently, activity of the *GIF* promoters and the localisation patterns of the encoded proteins was found to be linked to alterations in the cell cycle front, as delineated by the CYCB1;1 activity (Donnelly *et al.*, 1999; Kazama *et al.*, 2010). Enhancement of *AtAN3* expression was also sufficient to increase CYCB1;1 levels (leading to an increased cell division rate), and caused a delay in the arrest front movement towards the base of the leaf, thus prolonging the division phase (Vercruyssen *et al.*, 2014).

Interestingly, a translational fusion with the GUS reporter and an *in situ* localisation analysis, respectively, showed that *AtAN3* promoter activity and mRNA distribution are restricted to mesophyll cells (Wang *et al.*, 2010; Kawade *et al.*, 2010;

2013). However, AtAN3 proteins are also detected in the epidermal layers of the leaf, indicating that they are able to translocate via the plasmodesmata to clonally distinct cell layers (Kawade *et al.*, 2013). When this movement of *AtAN3* to the epidermis is blocked, leaves cannot grow properly, a feature that was demonstrated by the lack of complementation in an *an3* mutant expressing an immobile triple GFP-tagged *AtAN3* (Kawade *et al.*, 2013). On the other hand, targeted expression of AtAN3-GFP specifically in the L1 or in the subepidermal layer by using a PDF1 or CLV1 promoter, respectively, still provided a translocated AtAN3 into other cell layers, and could restore the mutant phenotype (Kawade *et al.*, 2013).

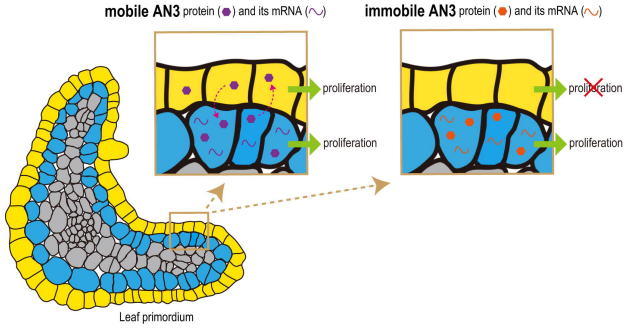


Figure 8. Movement of AN3 between cell layers coordinating proliferation between clonally distinct cells. *AN3* transcripts accumulate only in mesophyll cells, whereas AN3 proteins move from mesophyll to epidermal cells, leading to proliferation of the latter. Figure taken from (Kawade *et al.*, 2013)

These results thus indicate that AtAN3 carries out two functions with regard to the epidermis and the subepidermis: coordinator of cell proliferation within each layer, and signalling molecule for communication between them (Figure 8) (Kawade *et al.*, 2013; Hoe Kim and Tsukaya, 2015). It also provides an insight into how the epidermis can alter the cell division of mesophyll cells (and vice versa) and influence the final leaf size and shape (Marcotrigiano, 2010; Hoe Kim and Tsukaya, 2015). In maize, in which generating LOF mutants is more difficult due to a low transformation efficiency, the function of ZmAN3 and its redundant family members ZmGIF2 and ZmGIF3 is less well documented. However, as interaction studies of *AtAN3* and *ZmAN3* via TAP *in planta* showed a high level of conservation between the associating proteins, a similar role for *AN3* is suggested in both plant backgrounds. One of these conserved associations is the interaction between *AN3* and members of the GRF family of TFs, and another one is the binding of *AN3* with members of the SWI/SNF chromatin-remodelling complex (Vercruyssen *et al.*, 2014; Debernardi *et al.*, 2014; Nelissen *et al.*, 2015). Both are discussed below.

GIF+GRF: POSITIVE REGULATORS OF CELL PROLIFERATION

In *Arabidopsis*, *At*GRF expression is seen in several root and shoot parts, and is frequently associated with actively proliferating tissues (Kim *et al.*, 2003; Horiguchi *et al.*, 2005; Kim *et al.*, 2012; Liang *et al.*, 2014; Bao *et al.*, 2014). Analysis of plants carrying a mutation in the first three family members, *i.e. Atgrf1/2/3* triple mutants,

showed that their leaves are smaller and narrower (due to a reduction in mesophyll cell number), and have a shorter petiole. Remarkably, the phenotype of mutants that were defective in either *AtGRF1* or *AtGRF2* or *AtGRF3* was not altered. Overexpression of AtGRF1 and AtGRF2, on the other hand, results in larger leaves containing longer petioles. This indicates that these AtGRFs are functionally redundant, and have a regulatory function as enhancers of cell proliferation (Kim and Kende, 2004; Kim and Lee, 2006; Liu *et al.*, 2009). Studies on the other AtGRF members showed similar results, and measurements of all AtGRF transcript levels showed that they mostly accumulate in actively proliferating tissues during the initial phases of leaf development (Donnelly *et al.*, 1999; Rodriguez *et al.*, 2009). There, their expression patterns coincide with the expression pattern of markers of mitosis, such as the *CYCB1;1* gene (Donnelly *et al.*, 1999; Horiguchi *et al.*, 2005; Rodriguez *et al.*, 2009), a characteristic that seems to be conserved amongst eudicots and monocots (Yang *et al.*, 2009; Liu *et al.*, 2012; Baucher *et al.*, 2012). Besides their positive role in cell proliferation in the leaf, some AtGRFs, such as AtGRF4, are also involved in regulating the cell division during embryonic development of cotyledons and the SAM (Kim and Lee, 2006).

Although most of the AtGRFs have some redundant modes of action in regulating cell proliferation, the function of AtGRF5 seems to be very specific (Horiguchi *et al.*, 2005; Debernardi *et al.*, 2014). This is demonstrated by the fact that already a single-gene mutation in *Atgrf5* results in an altered phenotype, namely narrower leaves containing fewer cells (Horiguchi *et al.*, 2005). Moreover, overexpression of AtGRF5 generates an increased cell number and larger leaves, and was shown to stimulate chloroplast division, increased resistance against nitrogen depletion, and delayed leaf senescence (Horiguchi *et al.*, 2005; Vercruyssen *et al.*, 2015). Interestingly, the phenotypes of plants overexpressing AtGRF5 or AtAN3 are very similar. Y2H assays also showed direct interaction of both proteins, which indicates that both form a functional complex that enhances cell division and organ growth (Vercruyssen *et al.*, 2015).

The ZmGRF family in *Zea mays* is less well documented. Whereas generating LOF mutants in maize is difficult for large redundant protein families, such as the ZmGRFs, working with *Zea mays* leaves has the advantage that the developing tissues are larger, and that the growth zones are linearly organised along the length of the leaf (Nelissen *et al.*, 2012; 2015). This linear organisation allows the distinction between dividing and expanding tissues through high-resolution sampling along the growth zone, which is not possible in *Arabidopsis*, in which leaves are smaller than 1 mm at the time of transition between these two phases. A recent study in maize using TAP-MS on different developmental tissues for the identification of interaction partners showed that all *ZmGRFs*, with the exception of ZmGRF4 and ZmGRF10, associate specifically with *ZmAN3* in dividing tissues (Nelissen *et al.*, 2015). This suggests a conserved role of the GRFs in the positive regulation of cell proliferation during leaf growth in monocots and dicots. Indeed, overexpression of a miR396-resistant ZmGRF1 ($ZmGRF1^R$) resulted in longer leaves due to an increased number of dividing cells only (Wu *et al.*, 2014; Nelissen *et al.*,

2015). On the other hand, in contrast to the other ZmGRFs, ZmGRF4 and ZmGRF10 mRNAs lack a miR396a target site, which explains their stable presence also in expanding tissues (Candaele *et al.*, 2014). As mentioned above, the C terminus of ZmGRF10 is almost entirely absent, causing loss of the transcriptional activation capacity, whereas the interaction with ZmAN3 and ZmGIF2 is maintained. Strikingly, overexpression of *ZmGRF10* results in smaller leaves and plants due to a decrease in cell proliferation, indicating that this GRF works as a negative regulator of growth (Wu *et al.*, 2014). Based on these two results, a model was proposed in which ZmGRF1 and ZmGRF10 compete for binding to ZmAN3, where association with ZmGRF1 will lead to cell division, and binding of ZmGRF10 will provide an inactive complex, favouring cell expansion (Nelissen *et al.*, 2015).

In *Arabidopsis*, seven of the nine AtGRF family members (except AtGRF5 and AtGRF6) contain a target site in their WRC domain for miR396. During leaf growth, the amplitude of target AtGRF mRNA negatively influences the expression of the precursor (pri-miR396) and miR396 itself. As for ZmGRF1^{R}, cell proliferation and final leaf size in *Arabidopsis* plants were enhanced when endogenously expressing miR396-resistant versions of *AtGRF* (Rodriguez *et al.*, 2009; Wang *et al.*, 2010; Liang *et al.*, 2014; Debernardi *et al.*, 2014). This suggests that establishment of the miR396-GRF interaction limits the cell proliferation capacity of GRFs to allow fine-tuning of leaf growth (Hoe Kim and Tsukaya, 2015), a mechanism that is thus conserved in both monocots and dicots.

ESTABLISHING LEAF ADAXIAL/ABAXIAL POLARITY THROUGH GIF-GRF

Besides their role in the regulation of cell proliferation during leaf growth, AtGIF and AtGRF proteins have also been reported to participate in the establishment of adaxial/abaxial leaf polarity after leaf initiation in *Arabidopsis*. As mentioned before, the most important regulators of the adaxial/abaxial identity are AS1 and AS2, which will form a heterodimeric complex involved in the transcriptional activation of adaxially expressed genes (Martienssen *et al.*, 2000; Semiarti *et al.*, 2001; Guo *et al.*, 2008). When *as1* and *as2* mutants are combined with mutations in *AN3* or overexpression of *MIR396*, the adaxial defect phenotype is slightly enhanced, and leaves are small and needle- or lotus-like (Wang *et al.*, 2010; Mecchia *et al.*, 2013). Although a similar effect is seen in some extremely strong MIR396a overexpressor lines, it is completely absent in *an3* mutants and *MIR396*-overexpressing plants (Wang *et al.*, 2010; Mecchia *et al.*, 2013). Besides abaxialisation of leaf vasculature structures, transcript levels of some abaxially specific genes were slightly increased in *as2 35S:MIR396* and *as2 an3* double mutants, whereas adaxially expressed genes showed the opposite effect. In *as2 35S:MIR396*, for example, FIL/YAB1 expression was no longer restricted to the abaxial part of the leaf (Wang *et al.*, 2010). It is however not clear if the AtGIF-AtGRF duo affects adaxial/abaxial patterning directly, or if the changes are unspecific, as no significant changes in expression of both families in the two domains has been observed (Rodriguez *et al.*, 2009; Wang *et al.*, 2010; Horiguchi *et al.*, 2010). One possibility is that they influence the spatio-

temporal distribution of meristematic and differentiated cells necessary for correct adaxial/abaxial patterning (Hoe Kim and Tsukaya, 2015).

MOLECULAR MECHANISMS OF GIF AND GRF DURING LEAF GROWTH

In addition to its interaction with GRFs, AN3 was also shown to interact *in vivo* with core members of the SWI/SNF chromatin-remodelling complexes in both monocots and dicots, similar to its orthologue SYT in humans (Middeljans *et al.*, 2012; Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015). During the process of chromatin remodelling, a central ATPase-like BRM or BRG1 helps positioning the nucleosomes, and associates with other subunits, such as TFs and cofactors, to loosen the connections between DNA and histone octamers (Clapier and Cairns, 2009). This enhances the accessibility of *cis*-regulatory sequences for TFs or RNA polymerase II. Among the interacting subunits were two SWI/SNF ASSOCIATED PROTEINS73 (SWP73A and SWP73B), two SWI3 proteins (SWI3C and SWI3D), and two ACTIN RELATED PROTEINs (ARP4 and ARP7) (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015). Transcriptomic analysis of the maize leaf proteome showed how all non-plant-specific core components are absent from mature tissues (Facette *et al.*, 2013). TAP experiments on ZmAN3 in both dividing and expanding cells along the leaf growth zone (Nelissen *et al.*, 2015) on the other hand showed that the core components associate with ZmAN3 in these developmentally active tissues. Interestingly, two different ATPases (BRM and SLAYED) were identified in these TAPs, indicating that ZmAN3 associates with two different SWI/SNF complexes (Nelissen *et al.*, 2015). In *Arabidopsis,* Co-IP using AN3 as bait provided an additional ATPase, namely CHR12 (Debernardi *et al.*, 2014). This suggests that GIF proteins are involved in chromatin remodelling in the growth zone and as such might influence the expression of genes during leaf growth. Hence, the complex associates with TFs, such as the GRF family, although other TFs, such as WOX13, have also been reported to associate with AN3 in *Arabidopsis* seedlings (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015)(own data).

WHICH ARE THE TARGET GENES OF AN3?

In *Arabidopsis,* several transcriptomic studies have been carried out to gain insight into the molecular pathways of AN3 during the growth process, using AtAN3 LOF and GOF plants as well as plants overexpressing AtGRF5 (Gonzalez *et al.*, 2010; Vercruyssen *et al.*, 2014). Several functionally distinct groups of proteins were shown to be up- or down-regulated in these studies, and potential target genes were selected for further ChIP-qPCR analysis to confirm binding of the AN3-associated complex to their promoter regions. Among these targets, many TFs, such as HECATE1 (HEC1), CYTOKININ RESPONSE FACTOR2 (CRF2), CONSTANS-LIKE5 (COL5) and HOMEOBOX33 (HB33), as well as rDNA silencing inducing histone deacetylases HDT1 and HDT2, were identified (Vercruyssen *et al.*, 2014). Interestingly, AtAN3 was also shown to directly regulate its own promoter (Figure 9), and to influence the transcription of some AtGRFs, such as AtGRF6 and AtGRF9 (Ichihashi *et al.*, 2011). SWP73B, which associates with AtAN3, also binds the

promoter regions of *AtAN3*, *AtGRF3*, *AtGRF5*, *HEC1*, and *COL5*. Moreover, mutations in *BRM*, another interactor of AN3, were shown to affect some of the AtAN3 target genes (Vercruyssen *et al.*, 2014).

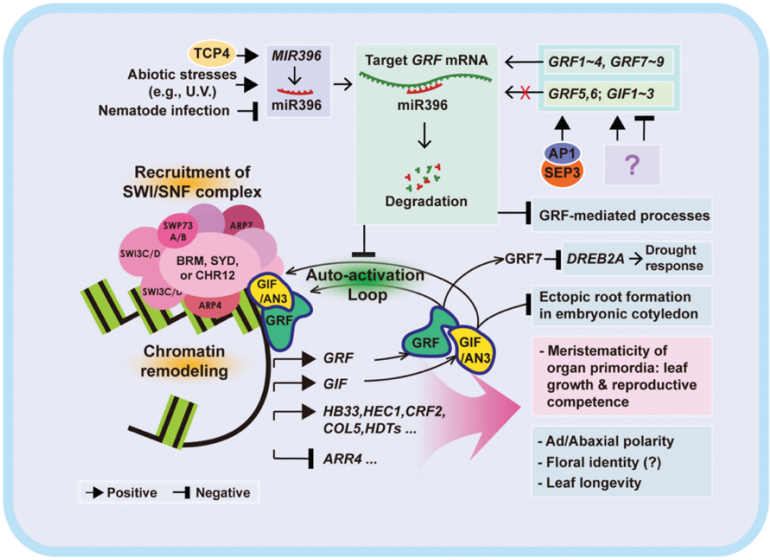


Figure 9. Schematic representation of the molecular and developmental roles of the GIF-GRF duo in *Arabidopsis*. Combinations of GIF and GRFs form complexes that regulate transcription during several biological processes. Active cell proliferation driven by GIF-GRF in organ primordia maintains their meristematic identities and results in successful reproduction and extensive leaf growth. Upstream of this is the recruitment of SWI/SNF chromatin-remodelling complexes to their target genes by AN3, which leads to activation or repression of their transcription by GRF. Target genes of GIF-GRF include *GRF* and *GIF* itself (creating an auto-activation loop), as well as many others, such as *HB33* and *ARR4*. Other factors that positively regulate *GRF* transcription are APETALA1 (AP1) and SEPALLATA3 (SEP3), suggesting a role for the duo in floral identity specification. One of the strongest regulators is miR396, which induces silencing of *GRFs* and thus negatively regulates the auto-activation loop. Figure taken from (Hoe Kim and Tsukaya, 2015).

In maize, RNA-seq-based transcriptome analysis on leaves of a *ZmGRF10*-overexpressing line provided a set of 176 up-regulated genes, whereas 74 genes were found to be down-regulated. Functional categories that were enriched among the up-regulated transcripts are mostly involved in metabolism or transcriptional regulation of development, such as SCARECROW-like, MADS-box TFs and homeobox genes; whereas the other group contained other TFs involved in growth/development, and some chromatin-modifying factors (Wu *et al.*, 2014). Although this suggests that ZmGRF10 is involved in various cellular mechanisms, further investigation is needed to unravel the molecular function of ZmGIFs and ZmGRFs in *Zea mays.*

Taking all of the above results together, the AN3 transcriptional co-activator might regulate cell proliferation during leaf growth by recruiting SWI/SNF complexes to the *cis*-elements that control genes involved in cell division, which

might influence their accessibility for GRFs or other stimulating or repressing TFs. In addition, GIF/GRF duos regulate their own expression level by means of a positive feedback loop, cooperatively with miR396.

REFERENCES

Achard P, Gusti A, Cheminant S, Alioua M, Dhondt S, Coppens F, Beemster GT, Genschik P. 2009. Gibberellin signaling controls cell proliferation rate in *Arabidopsis*.Current Biology **19**, 1188-93

Aizawa H, Hu SC, Bobb K, Balakrishnan K, Ince G, Gurevich I, Cowan M, Ghosh A. 2004. Dendrite Development Regulated by CREST, a Calcium-Regulated Transcriptional Activator. Science **303**, 197–202.

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell **121**, 207–221.

Andriankaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, De Milde L, Mühlenbock P, Skirycz A, Gonzalez N, Beemster GT, Inzé D. 2012. Exit from proliferation during leaf development in *Arabidopsis* thaliana: a not-so-gradual process. Developmental Cell **22**,64-78

Asl LK, Dhondt S, Boudolf V, Beemster GTS, Beeckman T, Inzé D, Govaerts W, De Veylder L. 2011. Model-based analysis of *Arabidopsis* leaf epidermal cells reveals distinct division and expansion patterns for pavement and guard cells. Plant Physiology **156**, 2172–2183

Autran D, Jonak C, Belcram K, Beemster GTS, Kronenberger J, Grandjean O, Inzé D, Traas J. 2002. Cell numbers and leaf development in *Arabidopsis*: a functional analysis of the STRUWWELPETER gene. The EMBO Journal **21**, 6036–6049.

Avramova V, Sprangers K, Beemster GTS. 2015. The Maize Leaf: Another Perspective on Growth Regulation. Trends in Plant Science **20**, 787–797.

Baloglu MC. 2014. Genome-wide in silico identification and comparison of Growth Regulating Factor (GRF) genes in Cucurbitaceae family. Plant Omics **7**, 260.

Bao M, Bian H, Zha Y, Li F, Sun Y, Bai B, Chen Z, Wang J, Zhu M, Han N. 2014. miR396a-Mediated Basic Helix-Loop-Helix Transcription Factor bHLH74 Repression Acts as a Regulator for Root Growth in *Arabidopsis* Seedlings. Plant & Cell Physiology **55**, 1343–1353.

Bao Z, Hua J. 2014. Interaction of CPR5 with Cell Cycle Regulators UVI4 and OSD1 in *Arabidopsis*. PLoS ONE **9**, e100347.

Baucher M, Moussawi J, Vandeputte OM, Monteyne D, Mol A, Pérez-Morga D, Jaziri El M. 2012. A role for the miR396/GRF network in specification of organ type during flower development, as supported by ectopic expression of Populus trichocarpa miR396c in transgenic tobacco (B Piechulla, Ed.). Plant Biology **15**, 892–898.

Beemster GTS, Vercruysse S, De Veylder L, Kuiper M, Inzé D. 2006. The *Arabidopsis* leaf as a model system for investigating the role of cell cycle regulation in organ growth. Journal of Plant Reseach **119**, 43-50

Bergmann DC, Sack FD. 2007. Stomatal development. Annual Review of Plant Biology **58**, 163–181.

Bonaccorso O, Lee JE, Puah L, Scutt CP, Golz JF. 2012. FILAMENTOUS FLOWER controls lateral organ development by acting as both an activator and a repressor. BMC Plant Biology 212, 176

Bowman JL, Eshed Y. 2000. Formation and maintenance of the shoot apical meristem. Trends in Plant Science **5**, 110–115.

Brett D, Whitehouse S, Antonson P, Shipley J, Cooper C, Goodwin G. 1997. The SYT protein involved in the t(X18) synovial sarcoma translocation is a transcriptional activator localised in nuclear bodies. Human Molecular Genetics **6**, 1559–1564.

Breuer C, Ishida T, Sugimoto K. 2010. Developmental control of endocycles and cell growth in plants. Current Opinion in Plant Biology, 1–7.

Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, Martienssen RA. 2000. Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. Nature **408**, 967–971.

Byrne ME, Simorowski J, Martienssen RA. 2002. ASYMMETRIC LEAVES1 reveals knox gene redundancy in *Arabidopsis*. Development **129**, 1957–1965.

Candaele J, Demuynck K, Mosoti D, Beemster GTS, Inzé D, Nelissen H. 2014. Differential methylation during maize leaf growth targets developmentally regulated genes. Plant Physiology **164**, 1350–1364.

Chitwood DH, Guo M, Nogueira FTS, Timmermans MCP. 2007. Establishing leaf polarity: the role of small RNAs and positional signals in the shoot apex. Development **134**, 813–823.

Cho HT and Cosgrove DJ. 2000. Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **97**, 9783–9788.

Choi D, Kim JH, Kende H. 2004. Whole Genome Analysis of the OsGRF Gene Family Encoding Plant-Specific Putative Transcription Activators in Rice (*Oryza sativa L.*). Plant & Cell Physiology **45**, 897–904.

Clapier CR, Cairns BR. 2009. The Biology of Chromatin Remodeling Complexes. Annual Review of Biochemistry **78**, 273–304.

Cosgrove DJ. 2005. Growth of the plant cell wall. Nat Rev Mol Cell Biol **6**, 850-61.

De Veylder L, Beeckman T, Beemster GTS, Krols L, Terras F, Landrieu I, Van Der Schueren E, Maes S, Naudts M, Inzé D. 2001. Functional Analysis of Cyclin-Dependent Kinase Inhibitors of *Arabidopsis*. The Plant Cell **13**, 1653.

Debernardi JM, Mecchia MA, Vercruyssen L, Smaczniak C, Kaufmann K, Inzé 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. The Plant Journal **79**, 413–426.

Deprost D, Yao L, Sormani R, Moreau M, Leterreux G, Nicolaï M, Bedu M, Robaglia C, Meyer C. 2007. The *Arabidopsis* TOR kinase links plant growth, yield, stress resistance and mRNA translation. EMBO reports **8**, 864–870.

Dill A. 2004. The Arabidopsis F-Box Protein SLEEPY1 Targets Gibberellin Signaling Repressors for Gibberellin-Induced Degradation. The Plant Cell **16**, 1392–1405.

Dill A, Jung HS, Sun TP. 2001. The DELLA motif is essential for gibberellin-induced degradation of RGA. Proceedings of the National Academy of Sciences **98**, 14162–14167.

Disch S, Anastasiou E, Sharma VK, Laux T, Fletcher JC, Lenhard M. 2006. The E3 Ubiquitin Ligase BIG BROTHER Controls *Arabidopsis* Organ Size in a Dosage-Dependent Manner. Current Biology **16**, 272–279.

Donnelly PM, Bonetta D, Tsukaya H, Dengler RE, Dengler NG. 1999. Cell Cycling and Cell Enlargement in Developing Leaves of *Arabidopsis*. Developmental Biology **215**, 407–419.

Douglas RN, Wiley D, Sarkar A, Springer N, Timmermans MCP, Scanlon MJ. 2010. ragged seedling2 Encodes an ARGONAUTE7-like protein required for mediolateral expansion, but not dorsiventrality, of maize leaves. The Plant Cell **22**, 1441–1451.

Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GTS, Inzé D, Ferreira PCG. 2011. The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. Plant Journal **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 GT, Remaut H, Van Montagu MC, De Jaeger G, Ferreira PC, Inzé 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.

Eshed Y, Izhaki A, Baum SF, Floyd SK, Bowman JL.. 2004. Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. Development **131**, 2997–3006.

Eshed Y, Baum SF, Perea JV, Bowman JL. 2001. Establishment of polarity in lateral organs of plants. Current Biology **11**, 1251–1260.

Facette MR, Shen Z, Björnsdóttir FR, Briggs SP, Smith LG. 2013. Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. Plant Cell **25**, 2798–2812.

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

Ferjani A, Horiguchi G, Yano S, Tsukaya H. 2007. Analysis of Leaf Development in fugu Mutants of *Arabidopsis* Reveals Three Compensation Modes That Modulate Cell Expansion in Determinate Organs. Plant Physiology **144**, 988–999.

Fisher K, Turner S. 2007. PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. Current Biology **17**, 1061–1066.

Filiz E, Koç İ, Tombuloğlu H. 2014. Genome-wide identification and analysis of growth regulating factor genes in *Brachypodium distachyon*: in silico approaches. Turkish Journal Of Biology **38**, 296–306.

Fournier C, Durand JL, Ljutovac S, Schäufele R, Gastal F, Andrieu B. 2005. A functional-structural model of elongation of the grass leaf and its relationships with the phyllochron. New Phytologist **166**, 881–894.

Geilfus C-M, Ober D, Eichacker LA, Mühling KH, Zörb C. 2015. Down-regulation of ZmEXPB6 (*Zea mays* β-expansin 6) protein is correlated with salt-mediated growth reduction in the leaves of *Z. mays* L. **290**, 11235–11245.

Geisler M, Nadeau J, Sack FD. 2000. Oriented asymmetric divisions that generate the stomatal spacing pattern in arabidopsis are disrupted by the too many mouths mutation. Plant Cell **12**, 2075–2086.

Gonzalez N, De Bodt S, Sulpice R, Jikumaru Y, Chae E, Dhondt S, Van Daele T, De Milde L, Weigel D, Kamiya Y, Stitt M, Beemster GT, Inzé D. 2010. Increased leaf size: different means to an end. Plant Physiology **153**, 1261–1279.

Gonzalez N, Pauwels L, Baekelandt A, De Milde L, Van Leene J, Besbrugge N, Heyndrickx KS, Cuéllar Pérez A, Durand AN, De Clercq R, Van De Slijke E, Vanden Bossche R, Eeckhout D, Gevaert K, Vandepoele K, De Jaeger G, Goossens A, Inzé D.2015. A repressor protein complex regulates leaf growth in *Arabidopsis*. The Plant Cell **27**, 2273–2287.

Gonzalez N, Vanhaeren H, Inzé D. 2012. Leaf size control: complex coordination of cell division and expansion. Trends in Plant Science **17**, 332–340.

Gordon SP, Chickarmane VS, Ohno C, Meyerowitz EM. 2009. Multiple feedback loops through cytokinin signaling control stem cell number within the *Arabidopsis* shoot meristem. Proceedings of the National Academy of Sciences **106**, 16529–16534.

Guenot B, Bayer E, Kierzkowski D, Smith RS, Mandel T, Zadnikova P, Benkova E, Kuhlemeier C. 2012. PIN1-

Independent Leaf Initiation in *Arabidopsis*. Plant Physiology **159**, 1501–1510.

Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, Cerwick SF, Dieter JA, Duncan KE, Howard RJ, Hou Z, Löffler CM, Cooper M, Simmons CR. 2013. Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. J Exp Bot **65**, 249–260.

Guo M, Thomas J, Collins G, Timmermans MCP. 2008. Direct Repression of KNOX Loci by the ASYMMETRIC LEAVES1 Complex of *Arabidopsis*. The Plant Cell Online **20**, 48–58.

Hay A, Barkoulas M, Tsiantis M. 2006. ASYMMETRIC LEAVES1 and auxin activities converge to repress BREVIPEDICELLUS expression and promote leaf development in *Arabidopsis*. Development **133**, 3955–3961.

Hershko A. 2005. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death and Differentiation **12**, 1191–1197.

Hoe Kim J, Tsukaya H. 2015. Regulation of plant growth and development by the GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR duo. J Exp Bot **66**, 6093–6107.

Hong S-Y, Kim O-K, Kim S-G, Yang M-S, Park CM. 2011. Nuclear import and DNA-binding of the ZHD5 transcription factor is modulated by a competitive peptide inhibitor in *Arabidopsis*. **286**, 1659–1668.

Horiguchi G, Kim G-T, Tsukaya H. 2005. The transcription factor AtGRF5 and the transcription co-activator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis* thaliana. The Plant Journal **43**, 68–78.

Horiguchi G, Nakayama H, Ishikawa N, Kubo M, Demura T, Fukuda H, Tsukaya H. 2010. ANGUSTIFOLIA3 Plays Roles in Adaxial/Abaxial Patterning and Growth in Leaf Morphogenesis. Plant & Cell Physiology **52**, 112–124.

Horiguchi G, Mollá-Morales A, Pérez-Pérez JM, Kojima K, Robles P, Ponce MR, Micol JL, Tsukaya H. 2011. Differential contributions of ribosomal protein genes to *Arabidopsis thaliana* leaf development. The Plant Journal **65,** 724-736

Howel SH. 1998. Molecular Genetics of Plant Development. Cambridge University Press.

Hu Y, Poh HM, Chua N-H. 2006. The *Arabidopsis* ARGOS-LIKE gene regulates cell expansion during organ growth. Plant Journal **47**, 1–9.

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

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

Inzé D, De Veylder L. 2006. Cell Cycle Regulation in Plant Development. Annual Review of Genetics **40**, 77–105.

Jackson D, Veit B, Hake S. 1994. Expression of maize KNOTTED1 related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development **120**, 405-413.

Jasinski S, Piazza P, Craft J, Hay A, Woolley L, Rieu I, Phillips A, Hedden P, Tsiantis M. 2005. KNOX Action in *Arabidopsis* Is Mediated by Coordinate Regulation of Cytokinin and Gibberellin Activities. Current Biology **15**, 1560–1565.

Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAs And Their Regulatory Roles In Plants. Annual Review of Plant Biology **57**, 19–53.

Juarez MT, Twigg RW, Timmermans MCP. 2004. Specification of adaxial cell fate during maize leaf development. Development **131**, 4533–4544.

Kalve S, De Vos D, Beemster GTS. 2014. Leaf development: a cellular perspective. Frontiers in Plant Science **5** 362

Kaplan DR. 1973. The Problem of Leaf Morphology and Evolution in the Monocotyledons. The Quarterly Review of Biology **48**, 437–457

Kato H, Tjernberg A, Zhang W, Krutchinsky AN, An W, Takeuchi T, Ohtsuki Y, Sugano S, de Bruijn DR, Chait BT, Roeder RG. 2001. SYT Associates with Human SNF/SWI Complexes and the C-terminal Region of Its Fusion Partner SSX1 Targets Histones. Journal of Biological Chemistry **277**, 5498–5505.

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

Kawade K, Horiguchi G, Usami T, Hirai MY, Tsukaya H. 2013. ANGUSTIFOLIA3 Signaling Coordinates Proliferation between Clonally Distinct Cells in Leaves. Current Biology **23**, 788–792.

Kazama T, Ichihashi Y, Murata S, Tsukaya H. 2010. The Mechanism of Cell Cycle Arrest Front Progression Explained by a KLUH/CYP78A5-dependent Mobile Growth Factor in Developing Leaves of Arabidopsis thaliana. Plant & Cell Physiology **51**, 1046–1054.**Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS**. 2001. Nature **411**, 706–709.

Kerstetter R, Vollbrecht E, Lowe B, Veit B, Yamaguchi J, Hake S. 1994. Sequence Analysis and Expression Patterns Divide the Maize knotted1-Like Homeobox Genes into Two Classes. The Plant Cell **6**, 1877.

Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. 2001. Nature **411**, 706–709.

Kim JH, Kende H. 2004. A transcriptional co-activator, AtGIF1, is involved in regulating leaf growth and morphology in *Arabidopsis*. Proceedings of the National Academy of Sciences **101**, 13374–13379.

Kim JH, Lee BH. 2006. GROWTH-REGULATING FACTOR4 of*Arabidopsis* thaliana is required for development of leaves, cotyledons, and shoot apical meristem. Journal of Plant Biology **49**, 463–468.

Kim JH, Choi D, Kende H. 2003. The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in*Arabidopsis*. The Plant Journal **36**, 94–104.

Kim JS, Mizoi J, Kidokoro S, Maruyama K, Nakajima J, Nakashima K, Mitsuda N, Takiguchi Y, Ohme-Takagi M, Kondou Y, Yoshizumi T, Matsui M, Shinozaki K, Yamaguchi-Shinozaki K. 2012. *Arabidopsis* GROWTH-

REGULATING FACTOR7 Functions as a Transcriptional Repressor of Abscisic Acid- and Osmotic Stress-Responsive Genes, Including DREB2A. The Plant Cell **24**, 3393–3405.

Kir G, Ye H, Nelissen H, Neelakandan AK, Kusnandar AS, Luo A, Inzé D, Sylvester AW, Yin Y, Becraft PW. 2015. RNA Interference Knockdown of BRASSINOSTEROID INSENSITIVE1 in Maize Reveals Novel Functions for Brassinosteroid Signaling in Controlling Plant Architecture. Plant Physiology **169**, 826–839.

Kondorosi E, Roudier F, Gendreau E. 2000. Plant cell-size control: growing by ploidy? Current Opinion in Plant Biology **3**, 488–492.

Krizek B. 2009. AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate *Arabidopsis* floral growth and patterning. Plant Physiology **150**, 1916–1929.

Kuijt SJH, Greco R, Agalou A, Shao J, 't Hoen CC, Overnäs E, Osnato M, Curiale S, Meynard D, van Gulik R, de Faria Maraschin S, Atallah M, de Kam RJ, Lamers GE, Guiderdoni E, Rossini L, Meijer AH, Ouwerkerk PB. 2014. Interaction between the GROWTH-REGULATING FACTOR and KNOTTED1-LIKE HOMEOBOX families of transcription factors. Plant Physiol **164**, 1952–1966.

Lee BH, Kim JH. 2014. Spatio-temporal distribution patterns of GRF-INTERACTING FACTOR expression and leaf size control. Plant Signaling & Behavior **9**, e29697.

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

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ, Turgeon R, Liu P, Sun Q, Nelson T, Brutnell TP. 2010. The developmental dynamics of the maize leaf transcriptome. Nature Genetics **42**, 1060–1067.

Li Y, Zheng L, Corke F, Smith C, Bevan MW. 2008. Control of final seed and organ size by the DA1 gene family in *Arabidopsis* thaliana. Genes & Development **22**, 1331–1336.

Liang G, He H, Li Y, Wang F, Yu D. 2014. Molecular mechanism of microRNA396 mediating pistil development in *Arabidopsis*. Plant Physiol **164**, 249–258.

Liu D, Song Y, Chen Z, Yu D. 2009. Ectopic expression of miR396 suppresses GRF target gene expression and alters leaf growth in *Arabidopsis*. Physiologia Plantarum **136**, 223–236.

Liu H, Guo S, Xu Y, Li C, Zhang Z, Zhang D, Xu S, Zhang C, Chong K. 2014. OsmiR396d-Regulated OsGRFs Function in Floral Organogenesis in Rice through Binding to Their Targets OsJMJ706 and OsCR4. Plant Physiology **165**, 160–174.

Liu J, Hua W, Yang H-L, Zhan G-M, Li R-J, Deng L-B, Wang X-F, Liu G-H, Wang H-Z. 2012. The BnGRF2 gene (GRF2-like gene from Brassica napus) enhances seed oil production through regulating cell number and plant photosynthesis. J Exp Bot **63**, 3727–3740.

Long JA, Moan EI, Medford JI, Barton MK. 1996. A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of *Arabidopsis*. Nature **379**, 66–69.

Marcotrigiano M. 2010. A role for leaf epidermis in the control of leaf size and the rate and extent of mesophyll cell division. American Journal of Botany **97**, 224–233.

Martienssen RA, Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A. 2000. Nature **408**, 967–971.

McConnell JR, Barton MK. 1998. Leaf polarity and meristem formation in *Arabidopsis*. Development **125**, 2935–2942.

McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK. 2001. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature **411**, 709–713.

Mecchia MA, Debernardi JM, Rodriguez RE, Schommer C, Palatnik JF. 2013. MicroRNA miR396 and RDR6 synergistically regulate leaf development. Mechanisms of Development **130**, 2–13.

Melaragno JE, Mehrotra B, Coleman AW. 1993. Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of *Arabidopsis*. The Plant Cell **5**, 1661–1668.

Middeljans E, Wan X, Jansen PW, Sharma V, Stunnenberg HG, Logie C. 2012. SS18 Together with Animal-Specific Factors Defines Human BAF-Type SWI/SNF Complexes (M Freitag, Ed.). PLoS ONE **7**, e33834.

Mizukami Y, Fischer RL. 2000. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proceedings of the National Academy of Sciences **97**, 942–947.

Nadeau JA, Sack FD. 2002. Control of Stomatal Distribution on the *Arabidopsis* Leaf Surface. Science **296**, 1697–1700.

Nakata M, Okada K. 2012. The three-domain model. Plant Signaling & Behavior **7**, 1423–1427.

Nakata M, Matsumoto N, Tsugeki R, Rikirsch E, Laux T, Okada K. 2012. Roles of the Middle Domain-Specific WUSCHEL-RELATED HOMEOBOX Genes in Early Development of Leaves in *Arabidopsis*. The Plant Cell Online **24**, 519–535.

Nath U, Crawford BC, Carpenter R, Coen E. 2003. Genetic Control of Surface Curvature. Science **299**, 1404–1407.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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..

Nelissen H, Gonzalez N, Inzé D. 2016. Leaf growth in dicots and monocots: so different yet so alike. Current Opinion in Plant Biology **33**, 72–76.

Nelissen H, Rymen B, Jikumaru Y, Demuynck K, Van Lijsebettens M, Kamiya Y, Inzé D, Beemster GTS. 2012. A Local Maximum in Gibberellin Levels Regulates Maize Leaf Growth by Spatial Control of Cell Division. Current Biology **22**, 1183–1187.

Nogueira FTS, Madi S, Chitwood DH, Juarez MT, Timmermans MCP. 2007. Two small regulatory RNAs establish opposing fates of a developmental axis. Genes & Development **21**, 750–755.

Oh M-H, Sun J, Oh DH, Zielinski RE, Clouse SD, Huber SC. 2011. Enhancing *Arabidopsis* Leaf Growth by Engineering the BRASSINOSTEROID INSENSITIVE1 Receptor Kinase. Plant Physiology **157**, 120–131.

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

Ori N, Cohen AR, Etzioni A, Brand A, Yanai O, Shleizer S, Menda N, Amsellem Z, Efroni I, Pekker I, Alvarez JP, Blum E, Zamir D, Eshed Y.2007. Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nature Genetics **39**, 787–791.

Osnato M, Stile MR, Wang Y, Meynard D, Curiale S, Guiderdoni E, Liu Y, Horner DS, Ouwerkerk PB, Pozzi C, Müller KJ, Salamini F, Rossini L. 2010. Cross Talk between the KNOX and Ethylene Pathways Is Mediated by Intron-Binding Transcription Factors in Barley. Plant Physiology **154**, 1616–1632.

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D. 2003. Control of leaf morphogenesis by microRNAs. Nature **425**, 257–263.

Pautler M, Eveland AL, LaRue T, Yang F, Weeks R, Lunde C, Je BI, Meeley R, Komatsu M, Vollbrecht E, Sakai H, Jackson D. 2015. FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize. The Plant Cell **27**, 104–120.

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. Genes & Development **18**, 2368–2379.

Peterson KM, Rychel AL, Torii KU. 2010. Out of the Mouths of Plants: The Molecular Basis of the Evolution and Diversity of Stomatal Development. The Plant Cell **22**, 296–306.

Rai MI, Wang X, Thibault DM, Kim HJ, Bombyk MM, Binder BM, Shakeel SN, Schaller GE. 2015**.** The ARGOS gene family functions in a negative feedback loop to desensitize plants to ethylene. BMC Plant Biology **15**, 157

Raventos D, Skriver K, Schlein M, Karnahl K, Rogers SW, Rogers JC, Mundy J. 1998. HRT, a Novel Zinc Finger, Transcriptional Repressor from Barley. Journal of Biological Chemistry **273**, 23313–23320.

Ren H, Santner A, del Pozo JC, Murray JAH, Estelle M. 2008. Degradation of the cyclin-dependent kinase inhibitor KRP1 is regulated by two different ubiquitin E3 ligases. The Plant Journal **53**, 705–716.

Rodriguez RE, Mecchia MA, Debernardi JM, Schommer C, Weigel D, Palatnik JF. 2010. Control of cell proliferation in *Arabidopsis* thaliana by microRNA miR396. Development **137**, 103–112.

Sakamoto T, Kamiya N, Ueguchi-Tanaka M, Iwahori S, Matsuoka M. 2001. KNOX homeodomain protein directly suppresses the expression of a gibberellin biosynthetic gene in the tobacco shoot apical meristem. Genes & Development **15**, 581–590.

Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. Genome biology **6**, 242.

Santos dos N, de Bruijn DR, Balemans M, Janssen B, Gärtner F, Lopes JM, de Leeuw B, Geurts van Kessel A. 1997. Nuclear localization of SYT, SSX and the synovial sarcoma-associated SYT-SSX fusion proteins. Human Molecular Genetics **6**, 1549–1558.

Scanlon MJ. 2003. The polar auxin transport inhibitor N-1-naphthylphthalamic acid disrupts leaf initiation, KNOX protein regulation, and formation of leaf margins in maize. Plant Physiology **133**, 597–605.

Scofield S, Murray JAH. 2006. KNOX Gene Function in Plant Stem Cell Niches. Plant Molecular Biology **60**, 929–946.

Semiarti E, Ueno Y, Iwakawa H, Tsukaya H, Machida C, Machida Y. 2001. The Asymmetric Leaves2 (AS2) Gene of *Arabidopsis* Thaliana Regulates Lamina Formation and is Required for Patterning of Leaf Venation. Progress in Biotechnology. Progress in Biotechnology, 63–68.

Shi J, Drummond B, Wang H, Archibald RL, Habben JE. 2016. Maize and Arabidopsis ARGOS proteins interact with ethylene receptor signaling complex, supporting a regulatory role for ARGOS in ethylene signal transduction. Plant Physiology, pp.00347.2016.

Shi J, Habben JE, Archibald RL, Drummonds BJ, Chamberlin MA, Williams RW, Lafitte HR, Weers BP. 2015**.** Overexpression of ARGOS Genes Modifies Plant Sensitivity to Ethylene, Leading to Improved Drought Tolerance in Both *Arabidopsis* and Maize. Plant Physiology. **169,** 266–282.

Sinha N, Hake S. 1992. Perturbations in leaf development caused by the dominant knotted-mutation in maize. Current Plant Science and Biotechnology in Agriculture. Progress in Plant Growth Regulation. Dordrecht: Springer Netherlands, 360–370.

Smith LG, Greene B, Veit B, Hake S. 1992. A dominant mutation in the maize homeobox gene, Knotted-1, causes its ectopic expression in leaf cells with altered fates. Development **116**, 21–30.

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. 2007.A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nature Genetics **39**, 623-630.

Spangenberg G. 2000. Molecular Genetics of Plant Development by Stephen H. Howell, Cambridge University Press. ISBN 0-521-587840-0; £24.95 (paperback). Plant Science **152**, 191.

Stransfeld L, Eriksson S, Adamski NM, Breuninger H, Lenhard M. 2010. KLUH/CYP78A5promotes organ growth

without affecting the size of the early primordium. Plant Signaling & Behavior **5**, 982–984.

Su YH, Liu YB, Zhang XS. 2011. Auxin–Cytokinin Interaction Regulates Meristem Development. Molecular Plant **4**, 616–625.

Su Z, Hao C, Wang L, Dong Y, Zhang X. 2011. Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum L.*). Theor Appl Genet **122**, 211–223.

Sugimoto-Shirasu K, Roberts K. 2003. 'Big it up': endoreduplication and cell-size control in plants. Current Opinion in Plant Biology **6**, 544–553.

Sun X, Cahill J, Van Hautegem T, Feys K, Whipple C, Novák O, Delbare S, Versteele C, Demuynck K, De Block J, Storme V, Claeys H, Van Lijsebettens M, Coussens G, Ljung K, De Vliegher A, Muszynski M, Inzé D, Nelissen H. 2017. Altered expression of maize PLASTOCHRON1 enhances biomass and seed yield by extending cell division duration. Nature Communications **8**, 14752.

Suzaki T, Sato M, Ashikari M, Miyoshi M, Nagato Y, Hirano H-Y. 2004. The gene FLORAL ORGAN NUMBER1 regulates floral meristem size in rice and encodes a leucine-rich repeat receptor kinase orthologous to *Arabidopsis* CLAVATA1. Development **131**, 5649–5657.

Taguchi-Shiobara F, Yuan Z, Hake S, Jackson D. 2001. The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. Genes & Development **15**, 2755–2766.

Thaete C, Brett D, Monaghan P, Whitehouse S, Rennie G, Rayner E, Cooper CS, Goodwin G. Functional domains of the SYT and SYT-SSX synovial sarcoma translocation proteins and co-localization with the SNF protein BRM in the nucleus. Human Molecular Biology **8**, 585-91

Thoma R, Chandler JW. 2015. Polarity in the early floral meristem of Arabidopsis. Plant Signaling & Behavior **10**, e992733.

Treich I, Cairns BR, de los Santos T, Brewster E, Carlson M. 1995. SNF11, a new component of the yeast SNF-SWI complex that interacts with a conserved region of SNF2. Molecular and Cellular Biology **15**, 4240–4248.

Tsukaya H. 2013*a*. Leaf Development. The *Arabidopsis* Book **11**, e0163.

Tsukaya H. 2013*b*. Does Ploidy Level Directly Control Cell Size? Counterevidence from *Arabidopsis* Genetics (GTS Beemster, Ed.). PLoS ONE **8**, e83729.

Tsukaya H, Uchimiya H. 1997. Genetic analyses of the formation of the serrated margin of leaf blades in *Arabidopsis*: combination of a mutational analysis of leaf morphogenesis with the characterization of a specific marker gene expressed in hydathodes and stipules. Molecular & general genetics : MGG **256**, 231–238.

van der Knaap E, Kim JH, Kende H. 2000. A Novel Gibberellin-Induced Gene from Rice and Its Potential Regulatory Role in Stem Growth. Plant Physiology **122**, 695–704.

Vanhaeren H, Nam Y-J, De Milde L, Chae E, Storme V, Weigel D, Gonzalez N, Inzé D. 2017. Forever Young: The Role of Ubiquitin Receptor DA1 and E3 Ligase BIG BROTHER in Controlling Leaf Growth and Development. Plant Physiology **173**, 1269–1282.

Veit B. 2008. Hormone mediated regulation of the shoot apical meristem. Plant Molecular Biology **69**, 397–408.

Vercruyssen L, Tognetti VB, Gonzalez N, Van Dingenen J, De Milde L, Bielach A, De Rycke R, Van Breusegem F, Inzé D. 2015. GROWTH REGULATING FACTOR5 Stimulates *Arabidopsis* Chloroplast Division, Photosynthesis, and Leaf Longevity. Plant Physiology **167**, 817–832.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L, Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. **26**, 210–229.

Veylder LD, Beeckman T, Beemster GTS, Krols L, Terras F, Landrieu I, Schueren EVD, Maes S, NAUDTS M, Inzé D. 2001. Functional Analysis of Cyclin-Dependent Kinase Inhibitors of Arabidopsis. The Plant Cell **13**, 1653.

Wang F, Qiu N, Ding Q, Li J, Zhang Y, Li H, Gao J. 2014. Genome-wide identification and analysis of the growth-regulating factor family in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). BMC Genomics **15**, 807.

Wang L, Gu X, Xu D, Wang W, Wang H, Zeng M, Chang Z, Huang H, Cui X. 2011. miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in *Arabidopsis*. J Exp Bot **62**, 761–773.

Wang X, Liu B, Huang C, Zhang X, Luo C, Cheng X, Yu R, Wu Z. 2012. Overexpression of Zmda1-1 gene increases seed mass of corn. African Journal Of Biotechnology **11**

Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmülling T. 2003. Cytokinin-Deficient Transgenic *Arabidopsis* Plants Show Multiple Developmental Alterations Indicating Opposite Functions of Cytokinins in the Regulation of Shoot and Root Meristem Activity. The Plant Cell Online **15**, 2532–2550.

White DWR. 2006. PEAPOD regulates lamina size and curvature in*Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **103**, 13238–13243.

Wu L, Zhang D, Xue M, Qian J, He Y, Wang S. 2014. Overexpression of the maize GRF10 , an endogenous truncated growth-regulating factor protein, leads to reduction in leaf size and plant height. Journal of Integrative Plant Biology **56**, 1053–1063.

Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, Gao F, Li Y. 2013. The Ubiquitin Receptor DA1 Interacts with

the E3 Ubiquitin Ligase DA2 to Regulate Seed and Organ Size in *Arabidopsis*. The Plant Cell **25**, 3347–3359.

Xie Z, Allen E, Wilken A, Carrington JC. 2005. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **102**, 12984–12989.

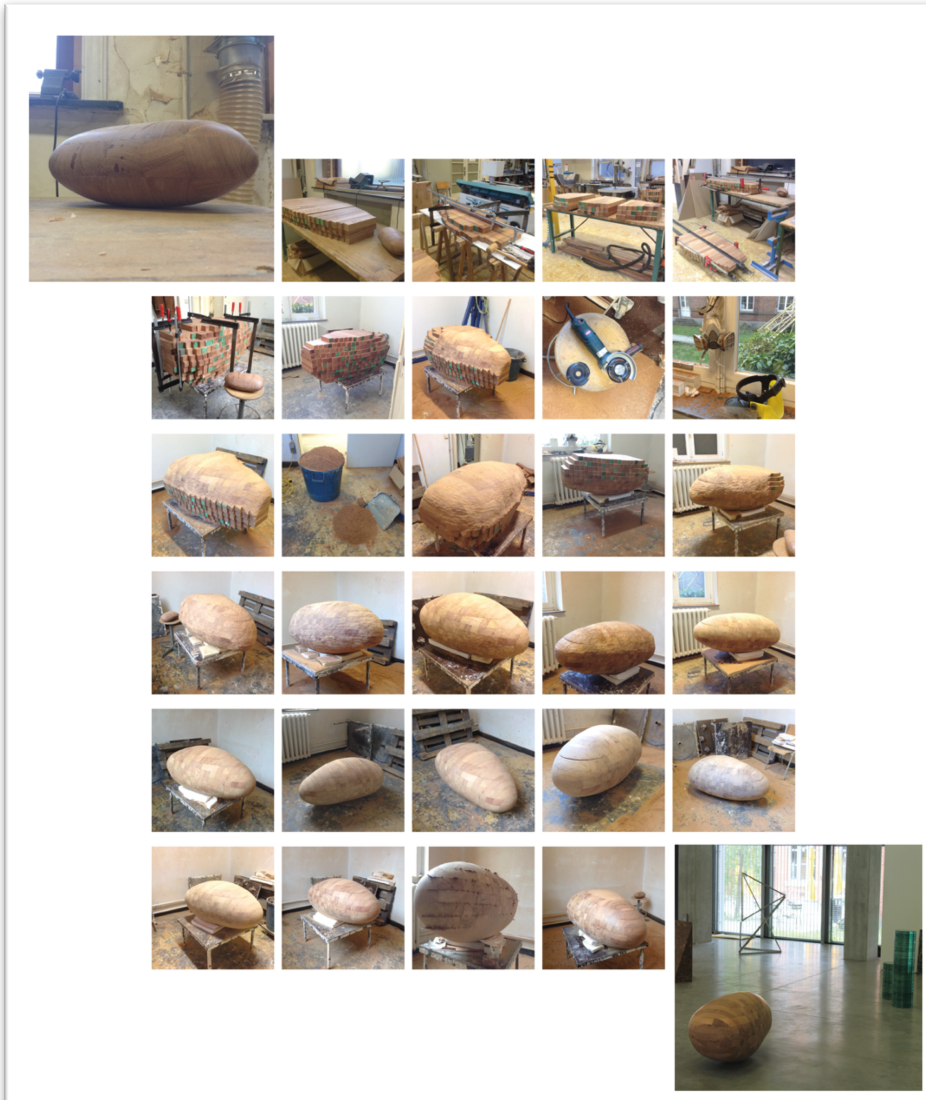
Yadav RK, Perales M, Gruel J, Girke T, Jonsson H, Reddy GV. 2011. WUSCHEL protein movement mediates stem cell homeostasis in the *Arabidopsis* shoot apex. Genes & Development **25**, 2025–2030.

Yanai O, Shani E, Dolezal K, Tarkowski P, Sablowski R, Sandberg G, Samach A, Ori N. 2005. *Arabidopsis* KNOXI Proteins Activate Cytokinin Biosynthesis. Current Biology **15**, 1566–1571.

Yang F, Bui HT, Pautler M, Llaca V, Johnston R, Lee B-H, Kolbe A, Sakai H, Jackson D. 2015. A maize glutaredoxin gene, abphyl2, regulates shoot meristem size and phyllotaxy. The Plant Cell **27**, 121–131.

Yang F, Liang G, Liu D, Yu D. 2009. *Arabidopsis* MiR396 Mediates the Development of Leaves and Flowers in Transgenic Tobacco. Journal of Plant Biology **52**, 475–481.

Zhang D-F, Li B, Jia G-Q, Zhang T-F, Dai J-R, Li J-S, Wang S-C. 2008. Isolation and characterization of genes encoding GRF transcription factors and GIF transcriptional co-activators in Maize (*Zea mays* L.). Plant Science **175**, 809–817.



Roden

Making Of – 2015-2016

Ciska L'Ecluse

Materials: Angelim pedra

GS^YELLOW/GREEN.: A MULTIFUNCTIONAL, "SWISS-KNIFE" TAP TAG FOR FUNCTIONAL PROTEIN ANALYSIS IN MONOCOTS AND DICOTS

Nienke Besbrugge[1,2], Bernard Cannoot[1,2], Dominique Eeckhout[1,2], Nancy De Winne[1,2], Geert Persiau[1,2], Eveline Van De Slijke[1,2], Daniel Van Damme[1,2], Mieke Van Lijsebettens[1,2], Shubhada Kulkarni[1,2], Klaas Vandepoele[1,2], Hilde Nelissen[1,2], Jelle Van Leene[1,2] and Geert De Jaeger[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2]Centre for Plant Systems Biology, VIB, 9052 Ghent, Belgium

(Manuscript in preparation for submission to Plant Physiology)

Author contributions

G.D.J., H.N. and J.V.L. designed the research; N.B. performed most of the experiments and wrote the article; D.E., N.D.W., E.V.D.S. and G.P. carried out experiments and provided technical assistance; D.V.D. and M.V.L. provided plant material; S.K. and K.V. helped with the bioinformatics analysis; J.V.L, H.N and G.D.J. supervised and complemented the writing

ABSTRACT

Tandem affinity purification coupled to mass spectrometry (TAP-MS) has become a standard approach to study the function of protein complexes in plants. We designed a highly efficient multifunctional TAP tag, called $GS^{green/}GS^{yellow}$, by combining the streptavidin-binding peptide (SBP) tag with a fluorescent protein (Green Fluorescent protein (GFP) or citrin Yellow Fluorescent Protein (cYFP)). Through a set of benchmark bait proteins, we show that this tag not only allows identification of interacting protein complex partners, but also visualisation of the bait, as well as identification of target genes from DNA-binding protein complexes through Chromatin Immunoprecipitation (ChIP)-seq, and this in *Arabidopsis thaliana* cell cultures and seedlings, and in leaves of *Zea mays*. Besides confirming previously reported data, the experimental data extend the current knowledge on the biological function of the baits. The $GS^{green/}GS^{yellow}$ TAP tag omits the need for the generation of distinct transgenic lines for three widely used molecular techniques, and thus provides a powerful tool for functional protein analysis in plants.

INTRODUCTION

Since the introduction of NGS methods, the availability of full genome sequences as well as technological advancements have driven the rise of the functional genomics field. Based on genome-wide and system-wide analyses of DNA sequences, gene expression and protein functions, this field attempts to describe the functions and interactions of genes and their products. By combining all gathered information into large molecular networks, an insight into complex biological processes, such as development and growth, can be obtained (Bunnik and Le Roch, 2013). Because proteins are the key players in these networks, their interactions with other proteins, DNA, RNA and metabolites are being extensively studied. In plants however, many genes and their products are still to be functionally annotated (Rhee and Mutwil, 2014).

To circumvent the need for specific antibodies for functional protein analysis, a genetically encoded affinity tag is often fused to the protein of interest, allowing affinity purification (AP) of the bait in its native context. Throughout the years, a variety of affinity tags have been developed each with their own specific characteristics enabling different functional analysis methods (Dedecker *et al.*, 2015). Some of them are short, such as the HA- or FLAG-epitope tags, some possess a very strong binding characteristic, such as peptide tags that get biotinylated *in vivo* and have a high affinity for (Strept)avidin ($Kd \approx 10^{-14}$), and some of them, such as GFP, can also function as a visual marker for localisation.

By combining AP with mass-spectrometry (AP-MS), individual proteins as well as protein complexes interacting with the bait protein can be identified. Two different AP-MS approaches are generally followed to study such PPIs, using either a single-step AP (also known as pull-down) or Tandem Affinity Purifications (TAP).

The TAP protocol, originally established in yeast (Rigaut *et al.*, 1999), is based on the addition of a dual affinity tag (originally named the TAP tag) to the bait of interest, allowing two subsequent rounds of protein purification. The high signal-to-noise ratios generated through this technique are favourable for the detection of sub-stoichiometric subunits, and provide very clean complexes for downstream biochemical analysis (of, for instance, enzymatic activity) (Dedecker *et al.*, 2015). In plants, the TAP-MS protocol has already been optimised several times, allowing purification of protein complexes from cell cultures (Van Leene *et al.*, 2014), whole plants/seedlings (Van Leene *et al.*, 2014) and isolated tissues (Nelissen *et al.*, 2015; Dedecker *et al.*, 2016), and across species, such as *Arabidopsis thaliana, Medicago truncatula*, *Zea mays* and *Oryza sativa* (Rohila *et al.*, 2006; Van Leene *et al.*, 2014; Nelissen *et al.*, 2015; Goossens *et al.*, 2016). Adaptation of the GS^{rhino} tag, an optimised version of the original TAP tag, even further enhanced the power of the protocol. This tag, consisting of two IgG-binding domains of protein G and a Streptavidin Binding Peptide (SBP) separated by two human rhinovirus 3C protease cleavage sites, was shown to outperform other TAP tags in plants (Van Leene *et al.*, 2011; 2014). Thanks to the recent advancements made in Label Free Quantification (LFQ) methods for quantitative protein analysis and the construction of sensitive MS detectors, TAP-MS can now even be used to determine the dynamics of protein complexes in plants, as was recently published for *Zea mays* AN3 (Nelissen *et al.*, 2015).

AP can also be used for genome-wide identification of target sites bound by transcription factors or other DNA-binding proteins by combining it with NGS methods. This technique, better known as ChIP-seq, starts by reversibly cross-linking the bait to its bound DNA regions *in vivo*. Next, nuclei are isolated and the chromatin is fragmented by for example sonication, rendering short bound DNA regions which are pulled down by a single (ChIP) affinity purification step. Next, the enriched genomic DNA is released by reverse cross-linking and identified through NGS. Another possibility is to perform a double (Tandem Chromatin Affinity Purification; TChAP) affinity purification, which does not require a nuclei isolation step due to the two purification rounds. In *Arabidopsis* cell suspension cultures, the implementation of this second purification step has been shown to enhance the DNA enrichment ratios significantly compared to ChIP approaches (Verkest *et al.*, 2014). The use of TChAP-seq for the identification of genome-wide target genes is indeed successful, as was recently exemplified for the transcriptional repressor PEAPOD2 (PPD2) (Gonzalez *et al.*, 2015) and for the TFs ERF115 (Heyman *et al.*, 2013) bZIP29 (Van Leene *et al.*, 2016), but biomass generation providing sufficient input can be challenging.

To minimise the production of transgenic lines for functional protein analysis, multifunctional tags that allow both visualisation of the bait and different AP based interaction studies have been constructed (Paramban *et al.*, 2004; Cheeseman and Desai, 2005; Kobayashi *et al.*, 2008; Schäffer *et al.*, 2010; Ma *et al.*, 2012; De Souza *et al.*, 2014). These multifunctional tags combine the power of fluorescent protein tags for visualisation (Giepmans, 2006; Voss *et al.*, 2013) with other tags that can be

used for single or tandem AP. Although their applicability across species, such as yeast, mammalian cells, *C. elegans, E. coli,* tobacco leaf extracts and *A. nidulans* has been reported (Paramban *et al.*, 2004; Cheeseman and Desai, 2005; Kobayashi *et al.*, 2008; Schäffer *et al.*, 2010; Ma *et al.*, 2012; De Souza *et al.*, 2014), successful implementation of such multifunctional tags in the identification of target genes by ChIP as well as their usage in the plant field is still missing.

Here, we describe the construction and evaluation of the $GS^{green/}GS^{yellow}$ TAP tag, a multifunctional TAP tag that allows both visualisation, ChIP, pull-down and TAP in plants from a single transgenic line. It combines elements from the highly efficient GS^{rhino} TAP tag with the visual GFP/cYFP protein tag that is often used in functional protein analysis in plants. Using a set of well-studied bait proteins involved in growth regulation as a benchmark, adaptation of the protocol and the performance of the $GS^{green/}GS^{yellow}$ tag in the identification of PPIs and target genes is shown in *Arabidopsis* cell cultures and seedlings, and in *Zea mays*. Dynamic changes in complex composition in different developmental stages could be captured. Overall, this new tag can provide a generic method for functional analysis of proteins and their interaction patterns in plants.

RESULTS

A PORTFOLIO OF MULTIFUNCTIONAL GS^{GREEN} AND GS^{YELLOW} TAGS

To broaden the functionality of the successful GS TAP tag in plants, the two IgG binding domains of protein G were replaced by a fluorescent protein tag, whereas the specific, high-affinity-binding Streptavidin-Binding Peptide (SBP) and the Gly4-Ser linker were retained. Two variants of this new multifunctional TAP tag were produced: GS^{green} and GS^{yellow}, containing an enhanced Green Fluorescent protein (eGFP) or an improved version of Yellow Fluorescent Protein (YFP), citrine YFP (cYFP) that has been used for tagging in *Zea mays* (Wu *et al.*, 2013), respectively (Figure 1). Besides a shift in the GFP excitation and emission peak to 514nm and 527nm, respectively, this cYFP thus shows an increase in brightness and resistance against photobleaching, acidic pH, and other environmental factors (Griesbeck *et al.*, 2001). Given the high similarity between GFP and cYFP (see Figure S1), the epitopes targeted in GFP-based affinity purification methods are conserved. Most experimental data reported here was gathered with GS^{yellow} tagged bait proteins.

Besides integration of a fluorescent protein in the tag, the GS^{green}/GS^{yellow} TAP tag also contains two copies of the Tobacco Etch Virus protease cleavage site and two of the human Rhinovirus 3C protease cleavage site. Given that the least efficient step in the TAP procedure is the protease-based elution step, the presence of two protease sites offers the opportunity to perform the specific elution at either 16°C (TEV) or 4°C (Rhino).

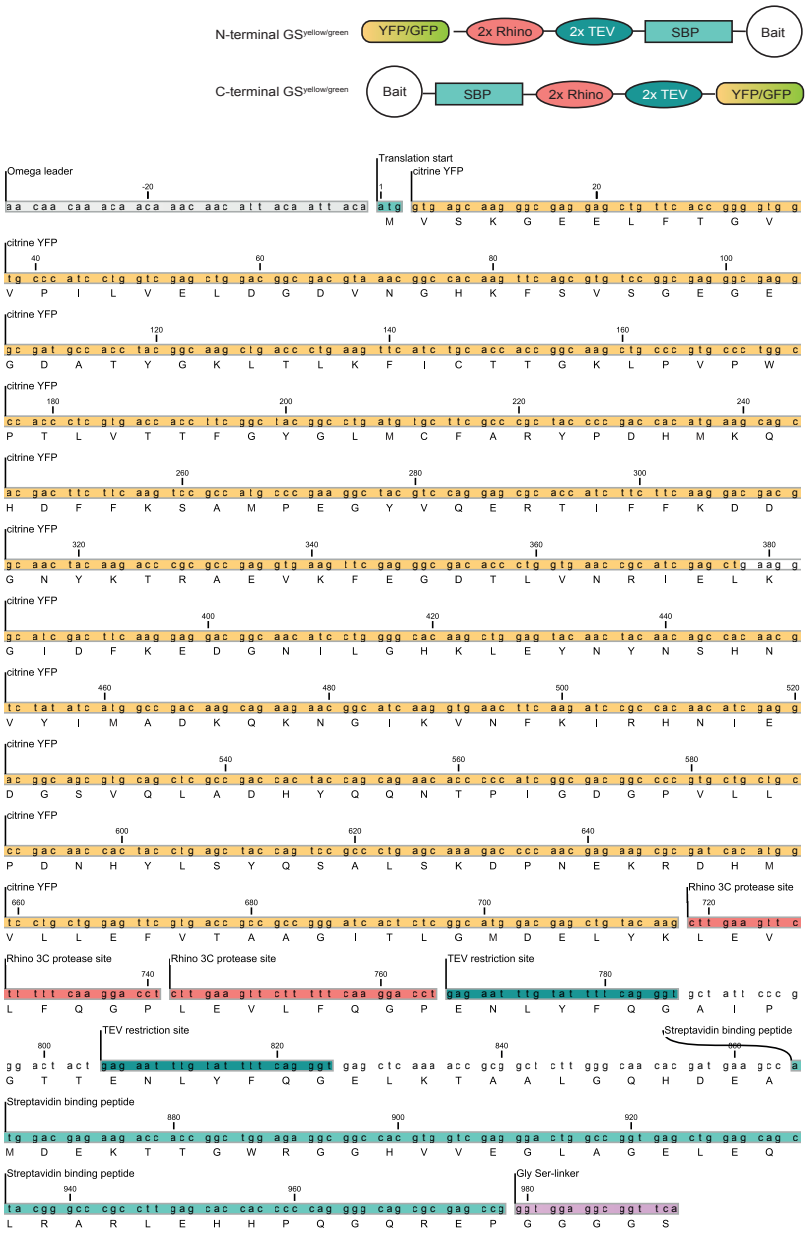


Figure 1. Schematic representation of the multifunctional GS$^{yellow/}$GS^{green} TAP tags. Top: Both C- and N-terminal fusions of the tag to the bait protein of interest are shown. YFP/GFP: yellow/green fluorescent protein; Rhino: cleavage site for the human 3C rhinovirus protease; TEV: cleavage site for the Tobacco Etch Virus protease; SBP: Streptavidin Binding Peptide. **Bottom:** Annotated DNA and amino acid sequence overview of the N-terminal GS^{yellow} tag

The resulting GS^{green}/GS^{yellow} tag has a molecular mass of 37,1 kDa, and has a broader applicability than previous tags (Table 1). Because N-terminal tagging sometimes interferes with correct subcellular localisation due to the presence of signal peptides at this terminus, an expansion of both the GS^{rhino} and GS^{green}/GS^{yellow} toolbox for N-terminal tagging was created by adding a signal peptide targeting the nucleus (NLS), the endoplasmic recticulum (ER), the chloroplast (Chl) or the mitochondria (Mito). To test whether these sequences result in correct localisation, a GFP protein was coupled at the N terminus to GS^{rhino} signal peptide TAP tags, showing the correct localisation of NGS^{rhino}-(NLS)-GFP, NGS^{rhino}-(ER)-GFP, NGS^{rhino}-(Chl)-GFP and NGS^{rhino}-(Mito)-GFP (Figure S2). Although the implemented ER signal encodes the 2S2 apoplast targeting sequence, it allows the visualisation of proteins that first will pass through the ER. In case the protein has to be retained at the latter location, a C-terminal KDEL sequence has to be added (Yun & Eipper, 1995). For efficient and flexible cloning, multisite Gateway-compatible ENTRY vectors for all versions of the GS^{rhino} and GS^{green}/GS^{yellow} are available, allowing either N- or C-terminal cloning to the gene of interest.

Table 1. Overview of tags and techniques commonly used for functional analysis of plant proteins

Tag	**Application**	**Molecular Mass**
GFP	Pull-down, ChIP, localisation	27.1 kDa
$GS^{green/}GS^{yellow}$	TAP, pull-down, ChIP, localisation	37.1 kDa
GS^{rhino}	TAP, pull-down	21.9 kDa
HBH	TChAP	9.63 kDa

Technique	**Product**	**Advantages**	**Disadvantages**
Pull-down	Protein complex	Quick Low input	High background levels Quantitative MS required so more experiments needed
TAP	Protein complex	High signal-to-noise Robust Downstream applications possible	Lower yield High input Mainly for stable interactions
ChIP	Target DNA	Low input → Tissue specification	Low DNA enrichment Limited reproducibility
TChAP	Target DNA	High signal-to-noise No need for nuclei isolation ID of less abundant interactions	High input Low yield Limited to cell cultures

TAP: Tandem Affinity Purification, ChIP: Chromatin Immunoprecipitation, TChAP: Tandem Chromatin Affinity purification

The performance of the GS^{green}/GS^{yellow} tag in DNA- or protein-interaction studies, as well as analysis of localisation patterns *in planta* is here evaluated. Hence, the bait proteins, AN3, TPLATE and PPD2, were selected based on previous publications of their interaction and/or localisation patterns in plant cell cultures and seedlings of *Arabidopsis thaliana*, or in *Zea mays*. Results of GS^{yellow}-tagged baits were compared to those reported in the literature.

ANALYSIS OF PROTEIN FUNCTIONALITY OF GS^{YELLOW} FUSION PROTEINS

Potential interference of a tag with the full functionality of the bait protein and/or its correct localisation is a known problem associated with protein tagging. An elegant way to test the functionality of the fusion protein, is expressing the tagged bait protein in the mutant background, and screening for complementation of the mutant. Another way to check for possible interference of the tag is by overexpressing (OE) the tagged protein of interest in a wild-type (WT) background, and assessing whether the OE phenotype is similar to the OE phenotype of its non-tagged counterpart.

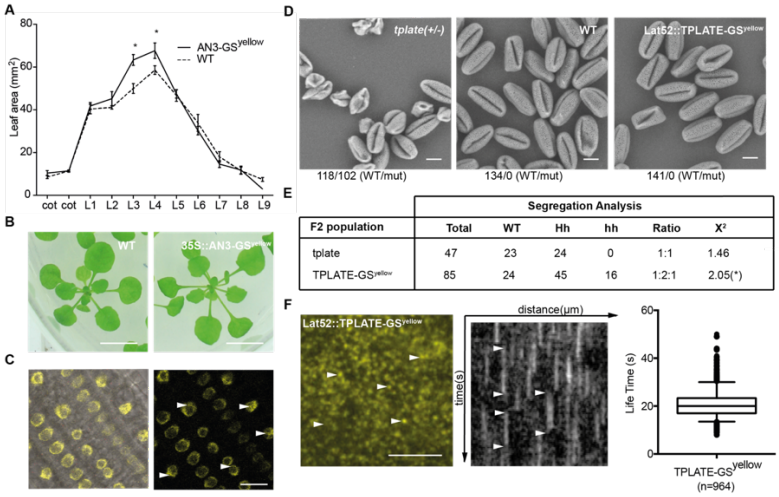


Figure 2. Functional analysis of GS^{yellow}-tagged AN3 and TPLATE. (A) Leaf area of 21-day-old WT or AN3- GS^{yellow} overexpressing plants grown on control medium. Cotyledons (cot) and leaves 1-9 were measured from leaf series. Error bars represent SE (n=3), asterisks indicate significant difference from wild type (*Col-0*)($p<0.05$, Student's *t*-test). **(B)** Representative rosettes of 21-day-old plants used for leaf series. Scale Bar = 1cm **(C)** Confocal microscopy picture of the root tip from 6-day-old seedlings expressing 35S:AN3-GSyellow. AN3 localises to the nucleus where it interacts with SWI/SNF complexes and forms speckles (indicated with arrowheads). Scale bar: 20µm. **(D)** SEM images of pollen grains coming from heterozygous *tplate* mutant plants, WT plants and heterozygous *tplate* mutants complemented with pLat52: TPLATE-GS^{yellow}. Ratios of WT versus mutant pollen are indicated. Scale bars = 10µm. **(E)** Segregation analysis of offspring produced by self-fertilisation of the heterozygous *tplate* mutant or the pLat52:TPLATE-GS^{yellow}-complemented mutant. Chi-square (X^2) values for the observed segregation ratios are indicated. Asterisks indicate significant difference from WT ratios ($p<0.05$) **(F) (left)** Representative spinning disc confocal microscope image of a hypocotyl epidermal cell belonging to plants complemented with pLat52:TPLATE-GS^{yellow}, taken from the corresponding movies. Discrete, dynamic foci represent recruitment of TPLATE, which can be measured over time through a kymograph (**middle**, arrowheads indicate beginning and end of the signal). TPLATE-GS^{yellow} has an average dwell-time of 20s **(right)** at the PM (n=934). Scale-bar = 5µm

To assess the functionality of GS^{yellow}-tagged proteins and their correct subcellular localisation *in planta*, two proteins, AN3 and TPLATE, were selected for further analysis in *Arabidopis* (Gadeyne *et al.*, 2014; Vercruyssen *et al.*, 2014). AN3 is a known plant growth regulator that functions as a transcriptional co-activator, linking the stable SWI/SNF chromatin-remodelling complex to TFs in both *Arabidopsis thaliana* and *Zea mays* (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015).

Previous experiments have shown that AN3 localises to the cell nucleus and forms speckles (Kim and Kende, 2004; Vercruyssen *et al.*, 2014), and that overexpression of AN3 in *Arabidopsis* enhances leaf growth due to a delay in exit from the mitotic cell cycle, leading to extended cell proliferation before transition to cell expansion. For functional analysis, AN3 was tagged at the C terminus with the GS^{yellow} TAP tag and expressed through the 35S promoter (p35S) in a WT (*Col*-0) background. At 21 days after stratification, overexpression of AN3-GS^{yellow} resulted in an increased leaf size and longer petiole phenotype of leaves 3 and 4 compared to WT plants (Figure 2). A positive effect, albeit not significant (Student's *t*-test, $p>0.05$), was also observed for leaves 1 and 2, which were reported in previous studies to be the most prevalently affected leaves by overexpression of AN3 (Horiguchi *et al.*, 2005; Vercruyssen *et al.*,2014). qRT-PCR showed that a lower level of AN3 overexpression in the AN3-GS^{yellow} line compared to the AN3 overexpression line that was studied before (Figure S3) might explain the differences in the severity of the observed growth phenotypes. In addition to the phenotypes, also the subcellular localisation of AN3 was confirmed, as the reported nuclear speckle localisation was observed in dividing root cells in the AN3-GS^{yellow} line (Figure 2) (Kim and Kende, 2004; Vercruyssen *et al.*, 2014).

The second benchmark protein is TPLATE, an adaptin-like protein involved in the early stages of clathrin-mediated endocytosis at the plasma membrane, where it dwells with an average lifetime of 20 s (Van Damme *et al.*, 2006; Gadeyne *et al.*, 2014). Plants lacking a functional *TPLATE* gene are male sterile due to problems with pollen germination and somatic cytokinesis, leading to shrivelled pollen. Reversal of this gametophyte-lethal phenotype can be achieved through expression of TPLATE from a pollen grain-specific promoter, pLat52 (Van Damme *et al.*, 2006). In our study, TPLATE was tagged at the C terminus with the GS^{yellow} TAP tag and expressed through the Lat52 promoter in a *tplate* mutant background. The fusion protein complemented the shrunken pollen grain phenotype (Figure 2) in three independent transgenic lines. Restoration of pollen germination was further confirmed by genotyping the offspring of heterozygous and complemented mutants through PCR. The progeny ratios shifted from the typical gametophyte-lethal 1:1 ratio to a normal mendelian inheritance pattern of 1:2:1 (χ^2: 1.8; $p<0.05$, Figure2) through complementation by the fusion protein. Moreover, the average dwell time of TPLATE-GS^{yellow} at the membrane was 20 s (n=964), which confirms the previously published dwell time (Gadeyne *et al.*, 2014). The above results indicate that the GS^{yellow} tag does not interfere with the functions of AN3 or TPLATE.

THE GS^{YELLOW} TAG PERFORMS EQUALLY WELL AS THE GS^{RHINO} TAG FOR THE ISOLATION OF PROTEIN COMPLEXES THROUGH TAP IN *ARABIDOPSIS*

Next, we wanted to study whether the multifunctional GS^{yellow} tag allows identification of protein interaction partners through TAP, and whether the performance of the GS^{yellow} tag in interactome mapping was similar to that of the

GS^{rhino} tag. AN3, TPLATE and PPD2 were selected as benchmark bait proteins for TAP in cell cultures. AN3 and TPLATE were also selected for TAP in seedlings because of the previous knowledge on their complex compositions in cell cultures and/or seedlings (Gadeyne *et al.*, 2014; Vercruyssen *et al.*, 2014; Gonzalez *et al.*, 2015). To compare the purification performance of the new GS^{yellow} TAP tag with the GS^{rhino} TAP tag, transgenic cell cultures with both GS^{rhino}- and GS^{yellow}-tagged versions of each bait protein were generated in parallel. After confirming accumulation of the fusion proteins in each culture through Western blot analysis, duplicate TAP experiments were performed using equal total protein inputs for each GS^{yellow}-GS^{rhino} bait pair, and purified components were detected by MS (see Supplemental Data S1A). To identify specific interactors for each bait, a list of non-specific proteins and contaminants was used to subtract false positive interactors from all proteins detected by MS. This list was constructed by performing control GS^{rhino} TAP experiments on extracts of WT cell cultures or seedlings, and combined with non-specific proteins detected across 803 TAP experiments with 163 different GS^{rhino} and two GS^{yellow} TAP-tagged fusion proteins (Van Leene *et al.*, 2014). After removal of these contaminants, a list of specific interactors found per tag is obtained for each bait protein (see Table 2).

Previously, a C-terminal fusion AN3-GS^{rhino} fusion was used for TAP in both *Arabidopsis* and *Zea mays*, showing that AN3 is part of the SWI/SNF chromatin-remodelling complex (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015) and associates with TFs, such as the GRFs and WUSCHEL-related homeobox 13 (WOX13) (Nelissen *et al.*, 2015). TAP on AN3-GS^{yellow} in cell cultures identified 13 of the 14 subunits of the SWI/SNF chromatin-remodelling complex, and four additional proteins (WOX13, unknown protein AT5G55210 and the two dentin sialophosphoprotein-related proteins AT5G07940 and AT5G07980) that are thought to either associate with the core SWI/SNF complex or with AN3 (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015) (see Table 2). For the GS^{rhino}-tagged version of AN3, only eight SWI/SNF core subunits were retrieved now, as well as the WOX13. When the TAPs with AN3-GS^{yellow} were performed on seedlings extracts, 15 out of the reported 16 interacting proteins found with TAP *in planta* (Vercruyssen *et al.*, 2014) were picked up through TAP (see Supplemental Data S1B). In addition, two extra core SWI/SNF complex subunits (BCL7B-1 (Nelissen *et al.*, 2015; Sarnowska *et al.*, 2016) and dentin salophosphoprotein-related AT5G07940) were also identified. Besides these 17 core proteins, the seedling experiment also affirmed the association of WOX13 that was found by the TAPs in cell cultures, as well as the interaction with GRF3 found in inflorescences (Debernardi *et al.*, 2014), albeit the latter was found only once in two experiments (see Table 2).

TPLATE, the second bait for evaluation, was reported to dynamically form a complex consisting of eight subunits involved in endocytosis (Gadeyne *et al.*, 2014). For the TPLATE cell culture TAPs, both the GS^{yellow} and GS^{rhino} TAPs provided the same list of interacting proteins, containing the reported octameric complex (Table 2). Subsequent TAPs on seedlings of the complemented pLat52:TPLATE-GS^{yellow}

transgenic lines further confirmed the association of TPLATE with the octameric complex in plants (see Table 2).

Table 2. Evaluation of GS^{yellow} performance in TAP experiments per bait in *Arabidopsis thaliana*

AN3		**Cell cultures***		**Seedlings ***	**Previously reported**
ProteinID	**Protein name**	**GS^{yellow}**	**GS^{rhino}**	**GS^{yellow}**	
AT5G28640	AN3 (bait)	2	2	2	x
AT1G18450	ARP4	2	1	2	x
AT3G60830	ARP7	2		2	x
AT4G22320	BCL7B-1			1	x
AT5G55210	BCL7B-2	2		2	x
AT1G20670	BRD1	2		2	x
AT5G55040	BRD13	2		1	x
AT1G76380	BRD2	2			x
AT2G46020	BRM	2	2	2	x
AT4G18480	CHLI1			2	
AT5G07940	dentin sialophosphoprotein-related	2		2	x
AT5G07980	dentin sialophosphoprotein-related	2			x
AT4G17330	G2484-1	2	2	2	x
AT5G17510	GLTSCR	2		2	x
AT2G36400	GRF3			1	x
AT3G22990	LFR			2	x
AT1G21700	SWI3C	2	1	2	x
AT4G34430	SWI3D	2	1	2	x
AT3G01890	SWP73A			2	x
AT5G14170	SWP73B	2	1	2	x
AT2G28290	SYD	2	2	2	x
AT4G35550	WOX13	1	1	2	X
TPLATE		**Cell cultures***		**Seedlings ****	**Previously reported**
ProteinID	**Protein name**	**GS^{yellow}**	**GS^{rhino}**	**GS^{yellow}**	
AT3G01780	TPLATE (bait)	2	2	2	x
AT3G50590	TWD40-1	2	2	2	x
AT5G24710	TWD40-2	2	2	2	x
AT5G57460	TML	2	2	2	x
AT2G07360	TASH3	2	2	2	x
AT1G20760	AtEH1	2	2	2	x
AT1G21630	AtEH2	2	2	2	x
AT1G15370	LOLITA	2	1	2	x
PPD2		**Cell cultures***			**Previously reported**
ProteinID	**Protein name**	**GS^{yellow}**	**GS^{rhino}**		
AT4G14720	PPD2 (bait)	2	2		x
AT4G28910	NINJA	2	2		x
AT4G32295	KIX8	2	2		x
AT3G17860	JAZ3	2	2		x
AT4G32570	TIFY8	2	1		x
AT3G24150	KIX9	1			x

Bait proteins are indicated in bold in left corner of each subdivision. Each list contains all the identified interaction partners. Constructs in cell cultures were expressed through the 35S constitutive promoter. For *in planta* TAPs (seedlings), we complemented for TPLATE the corresponding mutant with pLat52:TPLATE-GS^{yellow}, whereas for AN3 we transformed WT *Col-0* with 35S:AN3-GS^{yellow}. * Two technical repeats, ** Two biological repeats

Finally, the performance of the GS^{yellow} tag in TAP was analysed on PPD2. This protein was recently shown to regulate leaf development by repressing asymmetric

meristemoid division (Gonzalez *et al.*, 2015). TAP analysis in cell culture has shown that PPD2 interacts with proteins belonging to the TIFY class II and the KIX protein family (Gonzalez *et al.*, 2015). Here, the previously reported PPD2 interactors, NINJA and two proteins belonging to the TIFY protein family (jasmonate-ZIM-domain protein 3 and TIFY8) were identified with both TAP-tagged versions of the bait (see Table 2). Interactions with two KIX protein family members, KIX8 and KIX9, were identified with the GS^{yellow}-tagged bait protein but the latter could not be found in the interactor list of PPD2-GS^{rhino}.

Given the high overlap of the GS^{yellow} interaction data for three independent bait proteins with previously published data and with the GS^{rhino} TAP results, it is clear that the performance of the GS^{yellow} tag is comparable to the GS^{rhino} tag, allowing efficient protein complex analysis both in cell cultures and seedlings of *Arabidopsis.*

EXTENDING THE FUNCTIONALITY OF THE GS^{YELLOW} TAG FOR TARGET GENE IDENTIFICATION THROUGH CHIP IN *ARABIDOPSIS*

Besides implementing the GS^{yellow} TAP tag for the discovery of PPIs, a ChIP was performed to evaluate its performance in the identification of protein-DNA interactions. Previously, the GS^{yellow} tag was successfully applied for ChIP-qPCR confirmation of five PPD2 direct target genes (CYCD3;2, CYCD3;3, ALCATRAZ, SCHLAFMUTZE and PPD2). These were selected from a list of 13 target genes identified through integration of TChAP-seq data, obtained from a PPD2-HBH (His-Bio-His) expressing *Arabidopsis* cell culture, with transcriptome data from PPD2 knock-down plants (Gonzalez *et al.*, 2015). For the ChIP-qPCR confirmation, an existing GFP-based ChIP protocol (Morohashi *et al.*, 2012) was adapted for ChIP using GS^{yellow}.

In this study, a genome-wide ChIP-seq analysis was performed using PPD2-GS^{yellow} cell cultures to obtain a more detailed view on the performance of the GS^{yellow} tag in ChIP. ChIP-seq was done on two independent PPD2-GS^{yellow} experiments and on a nuclear localised version of the GS^{yellow} tag (NLS-GS^{yellow}) as negative control. After read mapping and peak calling, the identified targets were compared to data derived from two independent PPD2-HBH TChAP-seq experiments normalised to a single mock WT TChAP control. As the previously reported data set (Gonzalez *et al.*, 2015) of the first PPD2-HBH sample was created using an older version of the peak calling model, it was here reanalysed along with a new HBH TChAP and with the two GS^{yellow} ChIP experiments. Upon merging of the output of the two PPD2-HBH TChAP samples 1095 peaks corresponding to 1028 genes were identified (see Supplemental Data S2D-F), retaining 11 of the previously reported 13 direct candidate target genes in the overlap. For PPD2-GS^{yellow}, 1535 peaks corresponding to 1326 genes were found in common between both replicates (see Supplemental Data S2A-C).

Comparison of the genes obtained with the GS^{yellow} ChIPs with the 1028 genes shared in the overlap of the HBH TChAP repeats showed that 29% of all genes were found with both methods (see Figure 3). Moreover, 10 of the 11 target genes found

in the HBH TChAPs were also present in the GS^{yellow} dataset, demonstrating the relevance of the GS^{yellow} derived ChIP-seq data. Although the HBH tag provided one extra target, fold change enrichment values of the confirmed targets are much higher for the GS^{yellow} experiments. Here, five out of the ten confirmed targets were present in the top 100 of most enriched genes, compared to only one out of 11 for the HBH tag (Table 3). Both tag-specific datasets also show a significant enrichment for genes found to be differentially expressed in the Δ*ppd* deletion mutant (White).

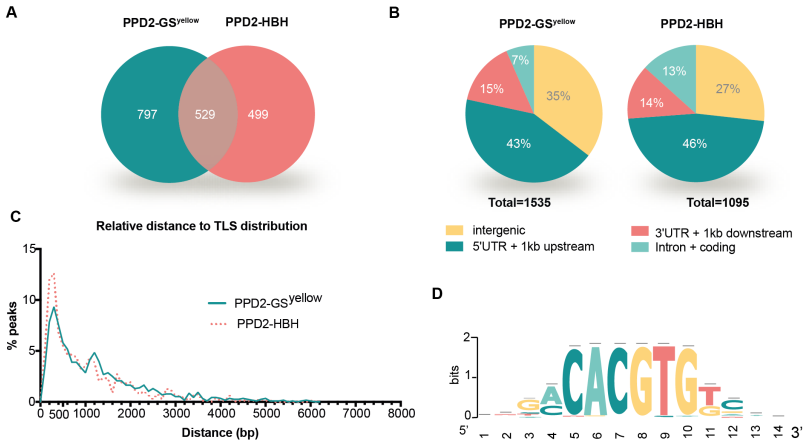


Figure 3. Genome wide analysis of PPD2-GS^{yellow} target genes through ChIP-seq. ChIP was performed with an anti-GFP antibody on cell suspension cultures (PSB-D) transformed with 35S:PPD2-GS^{yellow}. Results were compared with TChAP on PSB-D cultures transformed with 35S:PPD2-HBH (Gonzalez *et al.*, 2015) and normalised to a 35S:NLS-GS^{yellow}-expressing PSB-D culture. **(A)** Venn diagram of genes shared in each retrieved gene list after normalisation. **(B)** Genome-wide distribution of PPD2 DNA-binding sites in relation to the gene structure. Only peaks (identified with MACS2; Zhang *et al.*, 2008) present in both replicates were considered. **(C)** Distribution of the distance of peak summits to the nearest annotated Translation Start Site (TLS) in base pairs. **(D)** Graphic representation of the G-box motif bound by PPD2, as determined through *de novo* motif analysis

Analysis of the localisation of the mapped peaks retained by each tag demonstrates that most binding events occurred in the 5'UTR and 1 kb upstream regions (Figure 3), corresponding mainly to binding of PPD2 to gene promoter regions. Gene Ontology (GO) enrichment through the PLAZA platform (Proost *et al.*, 2015) for PPD2-GS^{yellow} also affirms the enrichment of genes involved in "DNA-dependent regulation of transcription" (p-value<1.53e-38) and "response to hormone" (p-value<8.42e-57)(see Figure S4 and Supplemental Data S3A-B). Moreover, the top 250 most enriched genes that were found with GS^{yellow} specifically segregate into similar GO categories as the GO of the genes found in the overlap (see Supplemental Data S3C-D). Motif analysis on the two separate PPD2-GS^{yellow} samples further confirmed overrepresentation (respectively 21% & 23% for the first and the second sample) of the reported G-box (CACGTG) (Gonzalez *et al.*, 2015) (see Figure 3).

Altogether, these data show that the multifunctional GS^{yellow} TAP tag can be used for ChIP-seq and ChIP-qPCR in *Arabidopsis* cell cultures. Moreover, based on

successful application of GFP-based ChIPs *in planta* (Kaufmann *et al.*, 2010; Yamaguchi *et al.*, 2014; Binkert *et al.*, 2016; Archacki *et al.*, 2016), one can expect that the GS^{yellow} tag can be used for ChIP in extracts obtained from *Arabidopsis* plants.

Table 3. Evaluation of GS^{yellow} performance in ChIP-seq through 35S:PPD2-GS^{yellow}-expressing *Arabidopsis* cell suspension cultures

		PPD2-GS^{yellow}		PPD2-HBH	
Closest gene	**Symbol**	**FE**	**-Log(q)**	**FE**	**-Log(q)**
AT5G49700	AHL17	***54.42***	***6.86***	6.12	2.66
AT5G67260	CYCD3;2	***53.78***	***7.56***	40.99	6.27
AT4G14720	PPD2	***51.29***	***4.89***	22.64	2.94
AT3G50070	CYCD3;3	***47.50***	***7.02***	35.11	6.30
AT1G14900	HMGA	***41.76***	***4.82***	27.53	3.61
AT3G54990	SMZ	38.92	6.28	30.08	4.81
AT2G37630	AS1	36.45	6.08	32.67	6.18
AT1G23870	TPS9	36.19	4.26	15.72	2.96
AT5G67110	ALC	34.83	5.87	42.00	6.84
AT5G54510	DFL1	27.07	5.13	***102.30***	***11.87***
AT5G59540	2OG-Fe(II)			5.43	2.70

Published target genes for PPD2 (Gonzalez *et al.*, 2015) and their Fold Enrichments (FE) in both the PPD2-GS^{yellow} and PPD2-HBH intersects. Bold underlined numbers indicate ranking of the FE value in the top 100 of highest enriched targets.

VALIDATION OF DYNAMIC PROTEIN-PROTEIN INTERACTIONS THROUGH TAP IN *ZEA MAYS* WITH THE GS^{YELLOW} TAP TAG

To evaluate the generic performance of the GS^{yellow} tag in other plant species, the multifunctionality of the tag was tested in *Zea mays*. The linear organisation of cell division and cell expansion along the *Zea mays* leaf allowed us to follow the dynamics of the ZmAN3 protein complex composition during leaf growth using TAP on ZmAN3-GS^{TEV}, the orthologue of AN3 in *Arabidopsis* (Nelissen *et al.*, 2015). In this way, several division zone-specific interactors, that are orthologues of the *Arabidopsis* GRFs, *i.e.*, TFs with a known growth-regulatory function in plants, were identified. The same study also showed that interactions of ZmAN3 with subunits of the SWI/SNF complex are conserved from dicots to monocots, and that these interactions last along the whole growth zone of the *Zea mays* leaf (Nelissen *et al.*, 2015). Using these results as a benchmark, the performance of the multifunctional GS^{yellow} tag in *Zea mays* was assessed by tagging ZmAN3 at the C terminus. To enhance the competition for interaction partners with the endogenous ZmAN3, a mildly constitutive ubiquitin promoter (pUBIL) was used. The resulting ZmAN3-GS^{yellow} line showed a slight alteration in leaf elongation rate (Supplemental Figure S5) but the overal final leaf size was the same as for nontransgenic siblings. Similar to the previously published ZmAN3-GS^{TEV} line, ZmAN3-GS^{yellow} plants developed more slowly than the nontransgenic siblings (Nelissen *et al.*, 2015). To confirm that the recombinant fusion protein was functional, we analysed its localisation pattern at the cellular level (see Figure 4). Confocal microscopy of the basal part of leaf 4 (*i.e.*, the division zone) showed accumulation of the YFP signal in the nucleus, where it clustered into speckles. This is similar to the localisation observed in *Arabidopsis* cells (Kim and Kende, 2004; Vercruyssen *et al.*, 2014).To evaluate the performance

of the GS^{yellow} tag in dynamic protein interactome mapping, we performed five TAP experiments coupled to LC-MS/MS on ZmAN3-GS^{yellow}. We sampled both the division and expansion zone of the emerging leaf 4. All samples were analysed on a Q-Exactive MS, and determination of the non-specific and sticky background proteins was carried out as previously published (Nelissen *et al.*, 2015). All identified proteins are presented in Supplemental Data S4A.

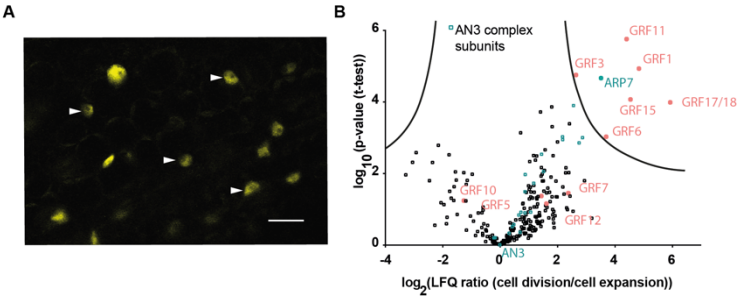


Figure 4. Functional analysis of ZmAN3-GS^{yellow} along the growth zone during leaf development in *Zea mays*. (A) Confocal microscopy picture of the division zone of leaf 4, 2 days after emergence. ZmAN3 localises to the nucleus and forms speckles (indicated by arrowheads). Scale bar = 20 µm. **(B)** Relative Intensity-Based Label-Free Quantification of co-purified proteins identified with TAP on ZmAN3-GS^{yellow} shown in a volcano plot. Comparison between the division zone and the expansion zone delineates significant changes of interaction partners between the zones. GRFs and AN3 core complex members are shown.

In short, 11 control TAP experiments were performed on different zones of leaf and ear tissues from non-transgenic mock plants to identify co-purifying contaminants, which are listed in Supplemental Data S4B. However, previous TAP experiments in *Arabidopsis* have shown that non-specific and sticky binders can better be identified using bait expressing cultures than mock samples (Van Leene *et al.*, 2014). Therefore, an additional filtering was applied on our dataset, excluding now all hits for which an orthologue was found to be present in an *Arabidopsis* TAP contaminant list that was constructed from 802 experiments on 163 different baits (Van Leene *et al.*, 2015, also provided in Supplemental Data S4C). The ZmAN3-GS^{yellow} TAPs mainly reiterated the previously identified dynamic AN3 interaction map (Table 4), and some additional co-purified proteins were present (Supplemental Table S1). Comparison of the GS^{yellow} results to the GS^{TEV} published data shows that we were able to co-purify and expand the complete SWI/SNF core complex with TAP in both growth zones. The dynamic nature of proteins belonging to the GRF family in the leaf growth zone was also mostly confirmed. Interestingly, GRF11, which was previously only found to bind to the SWI/SNF complex in the ear (Nelissen *et al.*, 2015), was also retained in leaf samples. Quantitative MS showed seven proteins, including six GRFs, to be significantly enriched in the division zone, whereas the bait, tag and co-purified core subunits did not change between both zones (Figure 4). The division zone-specific GRFs show a high overlap with the ones previously reported (Nelissen *et al.*, 2015). Taken together, these data show that the GS^{yellow} TAP tag performs as well as the GS^{TEV} TAP tag to isolate and map dynamic protein complexes in the leaf of *Zea mays*.

Table 4. Evaluation of GS^{yellow} performance in TAP experiments through ZmAN3 in *Zea mays*. The data were compared to previously published results in *Zea Mays* (Nelissen *et al.*, 2015).

ZmAN3		GS^{yellow}		Previously Reported	
Protein ID	**Protein Name**	**Division Zone(5)**	**Expansion Zone(5)**	**Division Zone(4)**	**Expansion Zone(4)**
Core Subunits					
GRMZM2G180246	AN3/GIF1	5	5	4	4
GRMZM2G015384	ARP4	5	5	4	4
GRMZM2G015861	ARP7	5	5	4	4
GRMZM2G018955	BCL7	5	5	4	2
GRMZM2G005646	BCL7	5	5		
GRMZM2G020548	BRD1	5	5	4	2
GRMZM2G044044	BRD1	5	5	2	1
GRMZM2G336962	BRD1	5	3		
GRMZM2G312501	BRD1			3	3
GRMZM2G353779	BRD1	5	5		
GRMZM2G163849	BRM	5	5	4	4
GRMZM2G103079	dentin sialophosphoprotein-related	5	5	4	4
GRMZM5G816791	dentin sialophosphoprotein-related	5	5	4	4
GRMZM2G473310	G2484-1	5	5	4	4
GRMZM2G033478	G2484-1	5	5	4	4
AC198518.3_FGP003	GLTSCR	5	5	4	4
GRMZM2G476652	LFR	5	5	3	3
GRMZM2G139760	SWI3C (SWIRM)	5	5	4	2
GRMZM2G340756	SWI3C (SWIRM)	5	5	4	4
GRMZM2G119261	SWI3D (CHB3)	5	4		
GRMZM2G047038	SWI3D (CHB3)	5	5	4	4
GRMZM2G052416	SWP73B (CHC1)	5	5	4	4
GRMZM2G467799	SYD	5	5	4	4
GRMZM2G387890	SYD	5	5	4	4
GRFs					
GRMZM2G034876	GRF1	5	3	4	2
GRMZM2G096709	GRF10	4	5	4	3
GRMZM2G067743	GRF11	5	4		
GRMZM2G119359	GRF12	5	4	1	
GRMZM2G178261	GRF15	(5*)		4	
GRMZM2G124566	GRF17	5	1	4	1
GRMZM2G045977	GRF18				
GRMZM2G105335	GRF3	(5*)		2	
GRMZM2G004619	GRF4			3	4
GRMZM2G129147	GRF5	5	1	2	
GRMZM2G041223	GRF6	5	1	3	
GRMZM5G850129	GRF7	5	4	4	

Number of experiments indicated between parentheses.
(*): Identified with one significant peptide

IDENTIFICATION OF ZMAN3 TARGET GENES THROUGH GS^{YELLOW}-BASED CHIP-SEQ IN *ZEA MAYS*

Finally, we assessed whether a GFP-based ChIP protocol using GS^{yellow} can also be applied to *Zea mays*. To this end, genome-wide target genes of AN3-GS^{yellow} were identified through ChIP-seq. Two replicate ChIP samples were generated from the growth zone encompassing the division, the transition zone and the expansion zone of the pUBIL:ZmAN3-GS^{yellow} transgenic line (see Figure 5). Two different negative controls were initially processed: (i) extracts derived from a *Zea mays* transgenic line expressing a nucleus-localised version of the GS^{yellow} tag

(pUBIL:ZmNLS-GS^{yellow}), and (ii) a 'No Antibody' Control (NoAB) where the anti-GFP antibody was omitted during the ChIP procedure on extracts derived from pUBIL:ZmAN3-GS^{yellow}. First of all, only peaks common between both ZmAN3-GS^{yellow} replicates were considered. Normalisation against the ZmNLS-GS^{yellow} control rendered 2158 peaks, mapping to 1838 genes, whereas with the NoAB control sample, 3130 peaks mapping to 2556 genes were identified (see Supplemental Data S5A-F). Both normalisation methods generated a similar peak distribution and peaks were predominantly situated in the intergenic and upstream regions, with a high fraction of peak summits found in the 1kb upstream region, reflecting many binding events in promoter regions (Figure 5). Although both control samples generate similar binding-site distributions, a lower number of genes was retained when using ZmNLS-GS^{yellow} as a negative control, indicating a more stringent filtering of non-specific immunoprecipitated DNA. Based on this observation, the use of the ZmNLS-GS^{yellow} negative control can be recommended for future ChIP-seq experiments in *Zea mays*.
To further increase the stringency of the obtained dataset, only peaks found in common between both normalisation methods were taken into account, narrowing down the peak list (1617 peaks) to 1421 genes. After sorting this dataset according to the fold enrichment ratio, many interesting targets with a possible role in leaf growth and development are amongst the most highly enriched genes. For example, the orthologue of the *Arabidopsis* genes AINTEGUMENTA (GRMZM2G399072), a key regulator of final leaf and flower sizes (Mizukami *et al.*, 1999, Krizek 2009), and of BARELY ANY MERISTEM3 (GRMZM2G052111), a receptor-like-kinase protein required for proper meristem function and organ development (DeYoung *et al.*,2006), were identified with a peak among the most significantly bound regions.

The binding appeared to be ZmAN3 specific (Figure 5) as it did not occur in the negative control samples. Many additional interesting target genes are present in the top 500 that can be linked to cell proliferation, such as ZmGRF3, ZmGRF9, ZmCYCD6;1 and ZmCYCD7;1, and members of the TCP TF family (Lima *et al.*, 2010; Buendía-Monreal *et al.*, 2011; Li, 2015; Nelissen *et al.*, 2015). It is interesting to observe that ZmAN3, the co-activator of GRFs, also binds to the promoter of ZmGRF3 and ZmGRF9, which could reflect a possible feedback mechanism, as is often seen with TFs that regulate their own expression as reported in *Arabidopsis* (Hoe Kim and Tsukaya, 2015). Further analysis of the network connectivity through integration into the STRING database (Franceschini et al., 2012) of the genes that have a peak in the top 500 revealed a significantly higher number of connected nodes than randomly would be expected (p-value 7,64e-12) (see Figure S7). When the *Arabidopsis* orthologues of the identified target genes were compared to the 2702 genes found in *Arabidopsis* (Vercruyssen et al., 2014), a highly significant (p-value <5.9e-96) overlap was found (see Supplemental Data S5G). Finally, a GO enrichment analysis using the PLAZA platform (Monocots v3.0, Proost et al., 2015), confirmed the previously reported regulatory function of ZmAN3 in leaf development and transcription regulation (see Figure 5 and Supplemental Data S5).

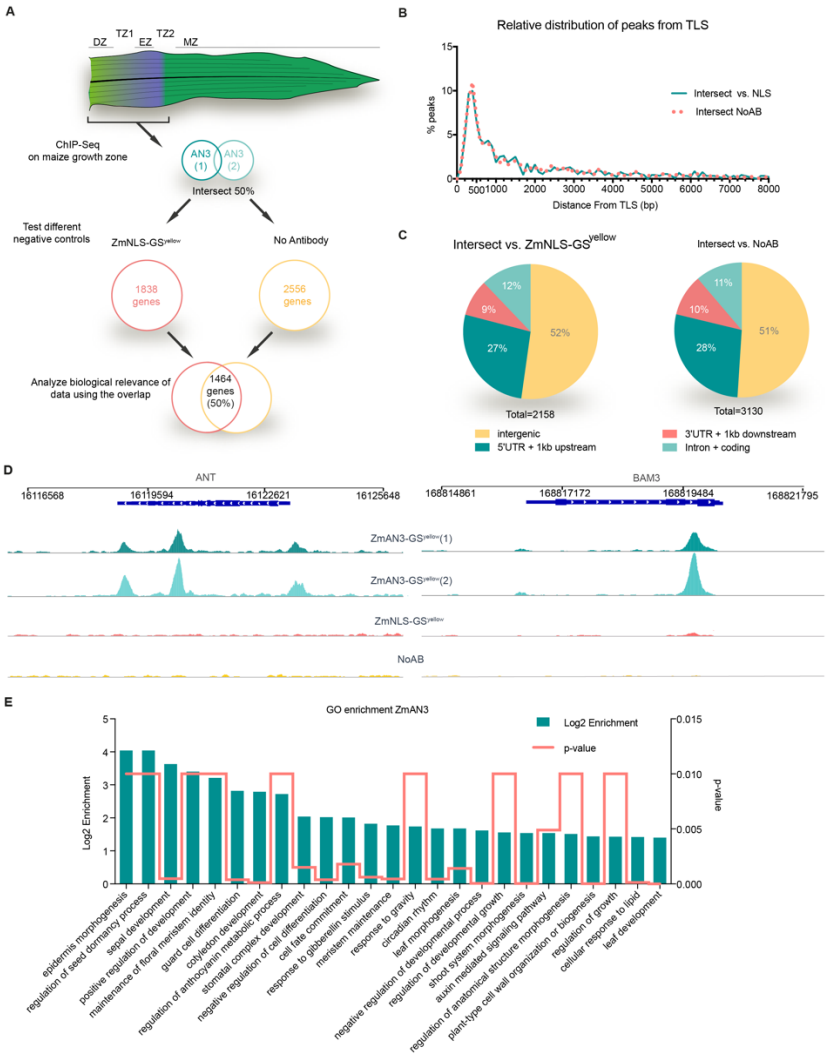


Figure 5. Identification of genome-wide ZmAN3-GS^{yellow} binding sites through ChIP-seq on the growth zone of leaf 4 in *Zea mays*. ChIP experiments were carried out in parallel on the full growth zone of leaf 4, 2 days after emergence. **(A)** Overview of the strategy used to identify genome-wide target genes of ZmAN3 through ChIP-seq. Normalisation against two negative controls (NoAB and ZmNLS-GS^{yellow}) was carried out on the 50% intersect of two ZmAN3 replicates. Final analysis was performed on the overlap of the genes present in each retained dataset. **(B)** Distance distribution per negative control in base pairs of ZmAN3 peak summits to the closest Translation Start Site (TLS). **(C)** Schematic representation of genome-wide ZmAN3 binding sites in relation to the gene structure. **(D)** Sashimi plot obtained from the Integrative Genomics Viewer. Visual presentation of ZmAN3, ZmNLS and NoAB sequence reads mapped to the transcription factor AINTEGUMENTA (ANT) and the receptor-like kinase BARELY ANY MERISTEM3 (BAM3), both involved in regulation of the leaf size. Gene regions are represented by blue bars. **(E)** Histogram of the GO enrichment calculated through PLAZA 3.0, Monocots (Proost *et al.*, 2012) on all genes present in the overlap of genes retained with each negative control. Categories with a log2 enrichment >1.4 and p<0.01 are shown.

It appears thus that not only the protein complexes in which ZmAN3 resides, but also its DNA-binding landscape is highly conserved from dicots to monocots. Taken together, these data show that ChIP-seq with the GS^{yellow} TAP tag in *Zea mays* was successful, and that the biological function of ZmAN3 is conserved among plant species.

DISCUSSION AND CONCLUSION

Functional analysis of proteins and their associated complexes through AP-MS is one of the most common methods to help unravel the complex molecular networks that regulate biological processes. To this end, different experimental techniques have been developed. As each of these techniques uses a variety of tags, correlated interpretation of the obtained results is hindered. In plants, protein complexes have been predominantly purified through TAP approaches (Dedecker *et al.*, 2015). We developed a multifunctional GS$^{green/}$GS^{yellow} TAP tag to allow the identification of both protein-protein and protein-DNA interactions, as well as gather information of the localisation pattern of the bait of interest from a single transformed line. Here we demonstrate that by combining fluorescent protein tags (GFP/cYFP) with the high affinity SBP tag in TAPs, very pure protein complexes could be purified from both cell cultures and seedlings of *Arabidopsis* as well as from *Zea mays* plants. The results confirmed most previously identified interactors of AN3, TPLATE and PPD2, and captured similar dynamic interaction patterns of ZmAN3-associated complexes during *Zea mays* leaf development. Although the GS^{yellow} here slightly outperformed the GS^{rhino} results for AN3 and PPD2 in cell cultures, the small variation observed between both tags seems rather bait- than tag-dependent (Van Leene *et al.*, 2008; 2014).

The multifunctionality of the TAP tag was further demonstrated by proving its usability for the identification of protein-DNA interactions through ChIP-seq in both cell cultures of *Arabidopsis* and leaves of *Zea mays*. The availability of GFP-based ChIP protocols for *Arabidopsis* plants (Gregis *et al.*, 2009; Kaufmann *et al.*, 2010) further suggests that the GS^{yellow} tag can also be used for ChIP *in planta*. The ChIP data using PPD2-GS^{yellow} cell cultures delivered higher enrichment levels of validated target genes than the HBH (TChAP) counterpart. Also, the genes specifically found with the GS^{yellow} tag show similar GO enrichments to those found by both tags, and are therefore estimated to be of higher biological value for further research. However, this apparent contradiction to previous results (Verkest *et al.*, 2014) might be explained by the difference in applied negative controls or expression levels of the baits, or could be due to the usage of new, strong ChIP-grade anti-GFP antibodies with magnetic beads in the GFP/cYFP-based ChIP protocol (Morohashi *et al.*, 2012; Gonzalez *et al.*, 2015). The strength of the tag for ChIP was also shown by the significant high overlap of target genes found with AN3 in *Arabidopsis* cell cultures and with ZmAN3 in the growth zone of developing *Zea mays* leaves. This demonstrates that not only the association of AN3 with SWI/SNF and GRFs is conserved between dicots and monocots (Nelissen *et al.*, 2015), but also the target genes of these complexes are to a certain extent conserved.

As the interaction with TFs, such as the GRFs was shown to be dynamic in nature along the growth zone (Nelissen *et al.*, 2015), it might be interesting to see whether this can also be captured at the DNA interaction level. Follow-up research could therefore focus on the mapping of dynamic gene-regulatory networks involving AN3/GRFs through differential ChIP on the different growth zones along the *Zea mays* leaf. Especially interesting here might be the availability of the CRISPR/Cas9 (Schaeffer and Nakata, 2015) genome editing tool for site-directed mutagenesis of *cis*-regulatory elements discovered by ChIP allowing targeted engineering of gene-regulatory networks in *Zea mays*.

Sometimes tagging can hamper the protein from forming interactions with its partner, losing its functionality (Palmer and Freeman, 2004). We showed here that coupling mutant complementation or overexpression analysis of the bait in *Arabidopsis* with predicted localisation data is a powerful tool to assess the functionality of the tagged protein. Although the phenotype of the AN3 overexpression was less pronounced in the first leaf pair, the growth increase was visible in leaves 3 and 4 suggesting at least partial complementation. In addition, the high overlap with previously published interaction and localisation data proves that the functionality of the tagged AN3 resembles its endogenous counterpart.

Here, genome editing methods may cause a major revolution in the field: on one hand, it allows the production of mutant backgrounds in species for which such mutant libraries are not yet available (such as *Zea mays*), and on the other hand it could be optimised for *in vivo* site-directed insertion of the tag sequences next to the gene of interest. The latter ensures that the endogenous landscape of genetic regulation is maintained, making complementation analysis an even more powerful tool to assess recombinant bait functioning.

In conclusion, the multifunctionality of the new GS^{yellow} TAP tag for functional protein analysis was shown through the identification of localisation patterns, protein-protein and protein-DNA interactions in both *Arabidopsis* and *Zea mays*. Performing these experiments with a single transgenic line allowed direct data comparison and elimination of positional effects of transgene insertion. This advantage makes the GS^{yellow} TAP tag a very attractive tool for usage with plant species that are difficult to transform, and emphasises the strength and usability of the tag for the plant research community.

MATERIALS AND METHODS

Construction of entry clones

The generation of the following entry clones through the Gateway recombination method (Invitrogen) has been described elsewhere: pENL1L2_AN3 (Vercruyssen *et al.*, 2014), pENL4R1_proLat52 and pENL1L2_TPLATE (Van Damme *et al.*, 2006), pENL1L2_PPD2 (White, 2006), pENL1L2_NLS and pENL4R1_proUBIL (Karimi *et al.*, 2007), and pENL1L2_ZmAN3 (Nelissen *et al.*, 2015). pENL1L2_ZmNLS (Varagona, 1992) was constructed through overlap PCR with

*attB1*_ZmNLS_Fwd and *attB2*_ZmNLS_Rev primers (Supplemental Table S3**)**, and cloned into pDONR221 via a BP reaction.

The GS^{green} and GS^{yellow} tags were constructed as follows. For the GS^{green}, two fragments, PacI_Ωleader_kozak_SBP-TEV2-Rh3C2-eGFP-G4S-XhoI for C-terminal and PacI_Ωleader_kozak_eGFP-TEV2-Rh3C2-SBP-G4S-XhoI for N-terminal tagging, were synthesised by GenScript in pUC57. The eGFP sequence was derived from pEGFPN1 (Clonetech). TEV2 and Rh3C2 sequences were copied from the GS^{TEV} (Van Leene *et al.*, 2011) and GS^{rhino} tag (Van Leene *et al.*, 2014), respectively. The omega leader sequence from tobacco mosaic virus was added to enhance translation. For the C-terminal GS^{green} tag, we used the vector with PacI_Ωleader_kozak_SBP-TEV2-Rh3C2-eGFP-G4S-XhoI as template in an adaptor PCR reaction to fuse attB2R_Gly4Ser-linker and attB3 to the GS^{green} tag sequence. This product-was then cloned into pDONRP2RP3 to construct pENR2L3_GS^{green}. For the N-terminal GS^{green} tag we used the PacI_Ωleader_kozak_eGFP-TEV2-Rh3C2-SBP-G4S-XhoI containing vector as template in an adaptor PCR reaction to fuse the *attB4/attB1R* or *attB1/attB2* Gateway sides to the GS^{green} tag sequence. The resulting PCR products were cloned into either pDONRP4P1R or pDONR221 to construct pENL4R1_NGS^{green} or pENL1L2_NGS^{green}.

For the C-terminal GS^{yellow} tag we first isolated cYFP DNA from DR5:PIN-cYFP *Zea mays* plants (Mohanty *et al.*, 2008), and amplified it using SBP+cYFP_Fwd and attB3_GFP/cYFP_Rev primers (listed in Supplemental Table S3). In parallel we amplified the SBP-TEV2-Rh3C2 fragment from the pENL1L2_GS^{green} using AttB2R_Gly4Ser_linker_Fwd and SBP+cYFP_Rev primers. Both PCR products were then combined using extension overlap PCR (Atanassov *et al.*, 2009) and cloned into pDONR221 to produce pENL1L2_GS^{yellow}. The N-terminal GS^{yellow} tag was synthesised in pUC57 as attB1-Ωleader-cYFP-Rh3C2-TEV2-SBP-G4S-*attB2* by Genscript, and cloned directly into pDONR221 using a BP reaction to produce pENL1L2_NGS^{yellow}. The pUC57 vector was then used as template for an adaptor PCR with *attB4/attB1R* Gateway sides. The resulting PCR product was finally cloned into pDONRP4P1R to produce pENL4L1R_NGS^{yellow}.

Construction of the signal peptide variants was carried out as follows: NLS (for nuclear targeting), mitochondrial (mito), chloroplast (chl), and 2S2 (for ER targeting) N-terminal GS^{rhino} and GS^{yellow} signal peptide TAP tag sequences were synthesised with minor codon adaptations and flanked by *attB1/attB2* gateway recombination sites. The sequences for the signal peptides were obtained from pENL1L2_NF (Karimi *et al.*, 2007), *Arabidopsis* 2S2 Albumin (Krebbers *et al.*, 1988), the E1 alpha subunit of plastid pyruvate dehydrogenase (Lee *et al.*, 2009) and the Cytochrome c oxidase subunit 4 (ScCox4) of *Saccharomyces cerevisiae* (Nelson *et al.*, 2007), respectively.

Synthesised fragments were used in standard Gateway BP reactions with pDONR221 to obtain *attL1/attL2* entry vectors. The resulting vectors were used as template for adaptor PCR and transfer of the tags to pDONRP4P1R.

Construction of expression vectors

The 35S:ORF-tag or 35S:tag-ORF constructs were generated as described by (Van Leene *et al.*, 2008). pLat52:TPLATE-GS^{yellow} was produced through a single Gateway LR reaction with the corresponding entry vectors and the pB7m34GW destination vector. PUBIL:ZmAN3-GS^{yellow} was constructed in the same manner, but with the pBb7m34GW destination vector (Karimi *et al.*, 2007).

Plant materials, growth conditions and growth measurements

PSB-D *Arabidopsis thaliana* cell suspension cultures were transformed with 35S:ORF-tag constructs, selected without callus screening and sub-cultured as described by (Van Leene *et al.*, 2008). Transformation of *Arabidopsis thaliana* (Columbia accession) plants was carried out through floral dip (Clough and Bent, 2008). 35S:AN3-GS^{yellow} was transformed into WT background and transgenic offspring was selected on ½MS medium containing kanamycin. *Arabidopsis* seedlings that expressed 35S:AN3-GS^{yellow} were grown in vitro for 6 days on ½ MS medium supplemented with 1% sucrose. PLat52:TPLATE-GS^{yellow} was transformed into the SALK_003086 line mutant background, which was provided by Daniel Van Damme, and transformants were selected on ½MS medium containing Basta. For complex analysis using TAP, *Arabidopsis* seedlings were grown as described by (Van Leene *et al.*, 2014). Analysis of the overexpressor phenotype of the AN3 lines was carried out as outlined by (Vercruyssen *et al.*, 2014). Briefly, plants were grown for 21 days in vitro on ½MS medium (Murashige and Skoog, 1962) supplemented with 1% (w/v) sucrose at 21°C under a 16-h-day/8-h-night regime. Leaf areas were measured with ImageJ (http://rsb.info.nih.gov/ij/) after dissection of individual leaves.

For *Zea mays* we transformed a B104 inbred line with the pUBIL:ZmAN3-GS^{yellow} via *Agrobacterium tumefaciens*-mediated transformation of immature embryos (Coussens *et al.*, 2012). This new transgenic line was then backcrossed to B104. The segregating offspring was selected using either AgraStrip® LL (PAT-assay) or fluorescent marker scoring through activity of pUBIL in dividing root tip tissue of emerging roots of sprouting seeds (Figure S4). Plants were grown and harvested as previously described (Nelissen *et al.*, 2015). For all transgenic plant lines, we selected lines that contained only a single locus for T-DNA insertion and showed a good level of protein expression on Western blots. Primary detection of GFP/cYFP tagged proteins was obtained with anti-GFP (*ab290,* Abcam©). Growth analyses were performed as described (Nelissen *et al.*, 2013).

Genotyping and analysis of phenotype complementation through SEM

Plant DNA was prepared from 1-2 rosette leaves as described by (Hsia *et al.*, 2010). Genotyping PCR of *tplate* (SALK_003086) lines was carried out as described by(Van Damme *et al.*, 2006). T-DNA and *pLat52:TPLATE-GS^{yellow}* positive plants were self-fertilized, and transmission of the T-DNA insertion on the offspring plants via the pollen was checked by PCR. Air-dried pollen coming from heterozygous

mutants transformed with *pLat52:TPLATE-GS*^{yellow} were tapped on stubs and examined with a Hitachi table top microscope TM-1000 (hitachi High-Technologies Europe BmbH, Germany), under a 15 kV acceleration voltage.

Tandem Affinity Purifications

TAP experiments on transformed cell suspension cultures were carried out as described by (Van Leene *et al.*, 2007). Briefly, per TAP, 2.5 g of biological material was ground in liquid nitrogen with mortar and pestle. The powder was then dissolved in 1.67 mL of ice-cold extraction buffer (BHB+: 25 mM Tris-HCl (pH 7.6), 15 mM $MgCl_2$, 150 mM NaCl, 15 mM p-nitrophenyl phosphate, 60 mM β-glycerophosphate, 0.1% (v/v) NP-40, 0.1 mM Na_3VO_4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTA-free Ultra complete tablet (1/10 ml) and 5% (v/v) ethylene glycol). After adding benzonase (0.1% v/v) and mixing with an Ultra-Turrax homogeniser, the mixture was incubated for 30 min at 4°C. Extracts were centrifuged twice, 20 min at 20,800g, 4°C. Supernatants were transferred to fresh tubes in between. The final supernatant was then filtered through a GF prefilter and a 0.45-µM syringe filter together. Protein concentrations were determined using Bradford assay and 25 mg full protein input extract was added to 25 µL equilibrated IgG-Sepharose beads (GS^{rhino}) or 25 µL anti-GFP agarose beads (GFP-Trap®_A, Chromotek, GS^{yellow}) and incubated for 1 h at 4°C on a tube rotator. Mixtures were filtered using a classic Mobicol column (ImTec) with a 0.35-µm bottom filter, and unbound fractions were removed through vacuum. Beads were washed five times with 750 µL. Wash buffer (100 µL) (10 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.1% (v/v) NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF and 5% (v/v) ethylene glycol) with 10 U Rhinovirus 3C protease were added to the Mobicol column. Supernatants were then incubated for 1 h at 4°C on a tube rotator, adding a second boost of 10 U after 30 min. Eluates were collected through centrifugation (1 min at 450g, 4°C), 100 µL extra wash buffer was added and collected through centrifugation. Next, the eluate was added to 25 µL equilibrated Streptavidin beads and incubated for 1 h at 4°C on a tube rotator. Mixtures were collected on a Mobicol column as before, and beads were washed three times with 833 µL wash Buffer. Finally, complexes were eluted by adding 30 µL of 1x NuPAGE lithium dodecyl sulfate sample buffer (Invitrogen) supplemented with 1x NuPAGE sample reducing agent (Invitrogen) and 20 mM desthiobiotin. The eluates were collected through centrifugation at 450g for 1 min at 4°C.

For the TAPs on *Arabidopsis* seedlings, the published protocol (Van Leene *et al.*, 2015) was used. This protocol is similar to the one used for cell cultures, but starts with 50 g ground plant material in liquid nitrogen to which 100 mL ice-cold extraction buffer was added, supplemented with 100 µL benzonase (0.1 % (v/v)). After incubation, tubes were centrifugated twice at 36,900g for 20 min at 4°C. Supernatants were filtered using a double layered Miracloth mesh. Equal full protein extracts were given using the maximum input possible in case of repeats to the same bead volumes. For the bead recovery steps, Poly-Prep chromatography columns (Bio-Rad) were used insteads of mobicol columns. After the second AP and

wash steps, complexes were eluted from the beads by applying 1 mL of streptavidin elution buffer (10 mM Tris/HCl, pH 7.6, 150 mM NaCl, 0.1% (v/v) NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% (v/v) ethylene glycol and 20 mM desthiobiotin) to the Poly-Prep column and incubating the beads for 5 min. Next, the eluate was collected by gravity in a 1.5-mL protein LoBind tube. Proteins were concentrated by adding 333 µL of trichloroacetic acid per 1 mL of streptavidin eluate (final concentration is 25% (v/v)), and the sample was mixed by inverting the tube. Proteins in the sample were precipitated over night through incubtaion of the sample on ice. Precipitated samples were then centrifuged in a microcentrifuge for 15 min at 20,800g at 4°C. The supernatants were removed, and the pellet was washed twice using 500 µL of an ice-cold solution containing 50 mM HCl in acetone. Between washes, samples were again centrifuged for 15 min at 20,800g at 4°C. The final supernatant was removed and the pellet was air dried for 1 h on ice. Samples were resolved with 30 µl of 1× NuPAGE sample buffer supplemented with 1× NuPAGE sample reducing agent.

The TAPs on *Zea mays* were executed as published (Nelissen *et al.*, 2015) with a protocol that is similar to the cell culture protocol of *Arabidopsis* but starting with 5 g biological material, adjusting the amount of buffer to the start material. Bead volumes and elution procedures were the same as for the cell culture samples.

In-Gel Protein Digestion

TAP eluates were heated for 10 min at 70°C and loaded onto a precast 4-12% (w/v) gradient NuPAGE Bis-Tris gel (Invitrogen). After running the gel for 7 min at 200 V, proteins were visualised through colloidal Coomassie Brilliant Blue G 250 staining. Destaining of the gel was carried out twice using 25 mL HPLC grade water for 1 h. Polypeptide disulfide bridges were reduced via incubation of the gel for 40 min in 25 mL of 6.66 mM DTT in 50 mM NH_4HCO_3. Subsequently, thiol groups were alkylated by incubating the gel in the dark for 30 min in 25 mL 55 mM iodoacetamide in 50 mM NH_4HCO_3 Next, the gel was washed by gentle shaking on an orbital shaker for 30 min in 25 mL of HPLC-grade water to remove the alkylating buffer. The gel was then transferred to a glass plate and a broad protein zone was cut out per lane using a fresh scalpel. Gel zones were sliced into 16-20 gel particles with a scalpel and transferred to a 1.5-mL protein LoBind tube. Gel particles were washed twice using 600 µL HPLC-grade water, and dehydrated in 600 µL 95% (v/v) acetonitrile for 10 min. The particles were then rehydrated twice in 600 µL HPLC-grade water for 10 min. Next, the particles were rehydrated for 30 min at 4°C in 90 mL digest buffer containing 1.125 mg trypsin (MS Gold; Promega), 50 mM NH_4HCO_3, and 10% (v/v) acetonitrile. Proteins were digested at 37°C for 3.5 h, and the trypsin solution was was transferred to a fresh 1.5-mL Protein LoBind tube. The trypsin sample was then incubated for 5 min in a sonicator bath. The gel particles were dehydrated again for 10 min with 300 mL 95% (v/v) acetonitrile. This acetonitrile solution was removed from the gel particles, and added tot he trypsin solution. Finally, the trypsin digest was dried in a cooled SpeedVac (2 to 3 h).

Liquid Chromatography-Tandem MS Analysis

The digested peptide mixtures were injected into a liquid chromatography-tandem MS system using an Ultimate 3000 RSLCnano LC (Thermo Fisher Scientific) in-line connected to either an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) (for the *Arabidopsis* samples) or a Q Exactive mass spectrometer (Thermo Fisher Scientific) (for the *Zea mays* samples). The sample mixture was first bound on an in-house prepared trapping column (100-mm i.d. 3 20 mm, 5-mm beads, C18 Reprosil-HD; Dr. Maisch). Next, the sample was flushed from the trapping column and loaded on an in-house prepared analytical column (75-mm i.d. 3 150 mm, 5-mm beads, C18 Reprosil-HD; Dr. Maisch) packed in the needle (PicoFrit SELF/P PicoTip emitter, PF360-75-15-N-5; New Objective). Loading of the peptides was carried out using a loading solvent consisting of 0.1% trifluoroacetic acid in water. Peptides were separated through a linear gradient from 98% solvent A9 (0.1% formic acid in water) to 40% solvent B9 (0.1% formic acid in water/acetonitrile, 20/80 [v/v]) for 30 min at a flow rate of 300 nL/min. Next, the column was washed for 5 min, hence reaching 99% solvent B9. The operation mode of the mass spectrometer was set at the data-dependent, positive ionisation mode, which automatically switches between MS and MS/MS acquisition for the 10 most abundant peaks in a given MS spectrum. For the LTQ-Orbitrap Velos, full-scan MS spectra were obtained in the Orbitrap set to a target value of 1E6 with a resolution of 60,000. The 10 ions with the highest intensities were then isolated and fragmented in the linear ion trap using a dynamic exclusion of 20 s. Fragmentation was carried out after filling the ion trap at a target value of $1E4$ ion counts. For the Q Exactive samples, the implemented source voltage was 3.4 kV and the temperature of the capillar was set at 275°C. After one MS1 scan (m/z 400 to 2000, AGC target $3\,3\,10^6$ ions, maximum ion injection time 80 ms) obtained at a resolution of 70,000 (at 200 m/z), up to 10 MS/MS scans (resolution 17,500 at 200 m/z) were obtained of the most intense ions that fitted the predefined selection criteria (AGC target $5\,3\,10^4$ ions, maximum ion injection time 60 ms, isolation window 2 D, fixed first mass 140 m/z, spectrum data type: centroid, underfill ratio 2%, intensity threshold $1.7\,3\,E4$, exclusion of unassigned, 1, 5 to 8, >8 charged precursors, peptide match preferred, exclude isotopes on, dynamic exclusion time 20 s). The HCD collision energy was kept at 25% of the normalized collision energy, and internal callibration (lock mass) was carried out using the polydimethylcyclosiloxane background ion at 445.120025 D.

Data Analysis and Background Filtering

For the *Arabidopsis* samples, a Mascot generic file was created using Mascot Distiller (version 2.4.3.1 for the cell culture samples, version 2.5 for the seedling samples). Spectra with a maximum intermediate retention time of 30 s and a maximum intermediate scan count of 5 were grouped, if possible, using a 0.005-Da precursor tolerance. De-isotoping was de-activated, and a relative signal-to-noise limit was set at 2. A peak list was generated only when the MS/MS spectrum contained more than 10 peaks. The peak list was then searched against the

TAIRplus database (accessible at http://www.psb.ugent.be/tapdata), which is an extended version of the TAIR10 database that also contains sequences of all types of non-TAIR contaminants in TAP or other proteomics experiments. Including these contaminants are the cRAP protein sequences that originate from proteins that were co-purified by accident or from contaminating proteins that are unavoidably present in protein samples (The Global Proteome Machine, http://www.thegpm.org/crap/). The searches were carried out using the Mascot search engine (version 2.4.1 for the cell cultures amples, version 2.5.1 for the seedling samples) with the following parameters: Variable modifications: Oxidation (methionine) and Methyl (aspartic acid); Fixed modification: Carbamidomethyl (cysteine); Enzyme: trypsin/proline; Maximum missed cleavages: 1; Charge: 2^+ and 3^+; Peptide tolerance: 10 ppm; # ^{13}C: 1; MS/MS tolerance: 0.5 Da; Instrument: ESI-TRAP. The Mascot data were then selected using a significance threshold of $p< 0.01$, with the maximum number of hits set to AUTO. Results were filtered in Excel for rank 1 peptides. Only identifications from *Arabidopsis* (TAIR10, this is the implemented database for all *Arabidopsis* searches) with at least two matched high-confidence rank 1 peptides, of which at least one is unique to the protein, were retained. The protein identification list was then cross-checked against a list of non-specific proteins (see below), and only specific binders were retained in the final list.

For the *Zea mays* samples, the raw files were processed in two sets (leaf versus controls) using MaxQuant software (version 1.4.1.2; Cox and Mann, 2008). The data were searched against the *Zea mays* database using the built-in Andromeda search engine. The latter database consists of all entries in the ZmB73_5a_WGS_translationsplus database (obtained through ftp://ftp.gramene.org/pub/gramene/maizesequence.org) and sequences of all possible types of contaminants that are often found with TAP or with proteomics experiments in general. These contaminants include common repository of adventitious protein sequences, a list with proteins that are often found in proteomics experiments, and are present as unavoidable contaminations of protein samples or by accident (The Global Proteome Machine, www.thegpm.org/crap/). In addition, tag sequences of many common tags and typical TAP contaminants, including sequences of resin proteins or implemented proteases, were added. The resulting ZmB73_5a_WGS_translationsplus database contains 136,770 sequence entries in total and can be accessed through www.psb.ugent.be/tapdata. For the searches, only oxidation of methionines and N acetylation of protein N termini were accepted as variable modifications, whereas fixed modifications were limited to carbamidomethylated cysteines. A mass tolerance for precursor (peptide) ions was set at 4.5 ppm, whereas for fragment ions this was set at 20 ppm. The used enzyme was trypsin/proline, permitting two missed cleavages as well as cleavage in cases where proline followed arginine or lysine. A complete list of the Max Quant search parameters is provided in Supplemental Table S4. A target-decoy approach with a False Discovery Rate (FDR) of 1% was used to filter protein identifications and PSM. Supplemental Data S4A contains the protein group output files from MaxQuant. Only proteins that were identified with at least two peptides, and of which at least one was unique, were retained (Table 4 and Supplemental Table S1).

Correcting for False Positives and Nonspecific Interactors

Besides specific, *bona fide* interactors, TAP experiments also co-purify non-specific interactors that interact with many unrelated bait proteins, as well as false positive interactors that bind the resin instead of the bait protein. Discriminating between such specific and non-specific or false positive interactors is generally carried out by performing a large series of TAP experiments with a wide variety of baits to determine bait-independent recurring co-purifying proteins (Keilhauer *et al.*, 2014; Van Leene *et al.*, 2014). For the *Arabidopsis* TAPs, we used an updated version of the previously published *Arabidopsis* background list (Van Leene *et al.*, 2014). This list was constructed by performing 803 TAP experiments on 163 different baits that were tagged with the GS tag and two witht the GS^{yellow} tag, and expressed in *Arabidopsis* cell cultures or seedlings. These baits were further subdivided into 87 groups, based on their co-occurence in the same protein complex (when this information was available) or on their involvement in the same biological process as assessed by their GO. For each protein that was identified in these experiments, a normalized spectral abundance factor (NSAF) was calculated by normalising the total number of spectra that identified the protein by its length and the total number of identified spectra in the sample. To avoid the loss of low-abundant *bona fide* interactors, a threshold of ΔNSAF $\geq$ 1% was set, so that the identified protein would not be classified as a non-specific binder, because the latter often have a minimal average NSAF value of 0.4% (Van Leene *et al.*, 2014). For each TAP experiment, possible interactors of the bait of interest were identified by filtering out all proteins that are present in this background list.

For the *Zea mays* TAPs, an extended version of a previously published *Zea mays* background list was used. This list was constructed by performing 11 mock GS^{TEV} purifications on wild-type B104 protein extracts obtained from leaf and ear, and extended with two mock GSyellow purifications on wild-type B104 protein extracts obtained from ear. From these mock purifications, all proteins that were identified with at least one high-confidence peptide with Q Exactive were used to build the extended maize background list (See Supplemental Data S4B). This list was then used to filter unspecific proteins from the TAP experiments on ZmAN3-GS^{yellow}. In addition, because the type of background in TAP experiments is determined to a large extent by the abundance of the unspecific proteins in the sample as well as by the type of purification that is implemented, we used the *Arabidopsis* background list that was used for the TAPs on *Arabidopsis* to further filter out unspecific background proteins in maize. Hence, we determined the *Arabidopsis* orthologues of all proteins that were identified in the maize TAP experiments via PLAZA 3.0 dicots (Proost *et al.*, 2015). If the orthologue of a protein in *Arabidopsis* was known to be in the background of Arabidopsis, the protein in maize was also considered an unspecific interactor and filtered out of the maize TAP data. All proteins that were identified in this manner are presented in Supplemental Data S4C. For future experiments, when more TAP data from a wider variety of baits will be available, implementing a maize-specific filtering using the frequency of occurrence of the co-purified proteins in a larger TAP dataset in maize can be applied (Van Leene *et al.*,

2014).

Label-Free Quantification (MaxLFQ) of Leaf ZmAN3 TAP Samples

For all ZmAN3-GS^{yellow} TAP samples, proteins in the samples were quantified in a relative manner using the MaxLFQ algorithm (Cox et al., 2014) that is integrated into the MaxQuant software. Therefore, only proteins identified with at least one unique peptide were used, and MaxLFQ values were determined only with unique peptides. Minimum LFQ ratio counts were set to 1, and the data were analysed with the Perseus software (version 1.6.0.2). Protein group files obtained from MaxQuant were loaded into Perseus, and all LFQ intensity values were logged. Reverse database hits and proteins only identified by site were filtered out.

All ZmAN3 TAP samples were grouped into division zone (five TAPs) or expansion zone (five TAPs). To improve confidence levels, proteins that did not show three valid values in at least one of both groups were removed. For missing values, a value was imputated based on a normal distribution around the detection limit of the mass spectrometer. We then performed a *t*-test to assess statistical outliers between the expansion and the division zone groups. To correct for the multiple hypothesis problem, a permutation-based FDR correction was used. The division zone-specific significantly enriched proteins identified with the *t*-test and the corresponding protein groups files can be found in respectively Supplemental Table S2 and Supplemental Data S4A.

Microscopy

Spinning disc confocal microscopy of endocytic foci and kymograph processing were carried out on hypocotyl cells as described (Dejonghe *et al.*, 2016). Time series were acquired for 1-2 min, with two time points per second. Excitation was done at 514 nm with detection between 550 and 615 nm. ImageJ (Fiji) was used for image processing. Background subtraction was achieved with a 5 pixel rolling ball radius, and walking averages of two subsequent frames were applied. Kymographs with a line-width of 3 pixels were made using the Kymograph2 plugin, and the lifetime of individual endocytic foci was measured by eye from the generated kymographs (n=964).

Transgenic *Zea mays* offspring were identified through the YFP signal in the root tip of emerging sprouts 1 week after sowing in a Jiffy-7® pellet. Screening for YFP fluorescence was carried out with a 1.0x lens on a MZ16 fluorescence stereomicroscope (Leica Microsystems) equipped with a blue light source and a Leica GFP Plus filter set. Images were obtained through the AxioCam HRc from Zeiss.

Confocal microscopy of 35:AN3-GS^{yellow} in *Arabidopsis* seedlings and of pUBIL:ZmAN3-GS^{yellow} in *Zea mays* was carried out with a FluoView1000 inverted confocal microscope (Olympus) equipped with a water-corrected 60x objective. cYFP was excited at 515 nm and its spectrum was exclusively detected between 530-560 nm. Background signal was identified as everything detected above

560 nm, which was sent to a different colour channel and then used as a negative filter for image acquisition.

Quantitative RT-PCR

RNA was extracted from 6-day-old seedlings expressing *35S:AN3-GS*^{yellow} using trizol combined with the RNeasy mini kit and on-column DNase digestion (Qiagen). cDNA was amplified from 1 µg of purified RNA with the reverse transcriptase containing iSCRIPT cDNA synthesis kit (Bio-Rad). qRT-PCR using the SYBR Green master mix (Roche) and gene-specific primers for *AN3* and *AN3-GS*^{yellow} was then performed. Two reference genes (*PP2AA3* and *CDKA1;1* (Hong *et al.*, 2010)) were selected for normalisation. Primers are listed in Supplemental Table S3.

ChIP-seq

ChIP-seq on PPD2-GS^{yellow} and ZmAN3-GS^{yellow} was performed with an anti-GFP antibody (*ab290,* Abcam©) according to an adapted protocol (Morohashi *et al.*, 2012). For each *Zea mays* sample, 1.2 g of tissue was fixed in Buffer A for 15 min under vacuum conditions. Glycine was added to a final concentration of 0.1 M, and incubation was continued for an additional 10 min. Fixed leaf tissue was washed with distilled water and frozen in liquid N_2. Cross-linking of *Arabidopsis* cell culture material was carried out as published (Verkest *et al.*, 2014), and 1 g of fixed material was used per ChIP. All subsequent steps were carried out as described (Morohashi *et al.*, 2012). Library preparation and sequencing of enriched DNA was done via Illumina sequencing on a Genome II Analyser (Illumina, (http://www.illumina.com/applications/sequencing.html). In total, 20 million reads were carried out per sample, with 50-bp-long single-end and 125-bp-long paired-end reads for *Arabidopsis* cell culture and *Zea mays* samples, respectively. The sequenced data were processed as described (Van Leene *et al.*, 2016) for the *Zea mays* samples. Here, peak calling was done with an effective genome size of 1920Mb and MFold parameters [10, 30]. Peaks were annotated against the ZmB73_5b.60 version of the genome, and shared peaks were identified using intersectBed with at least 50% overlap. Figure S8 gives and overview of the processing pipeline (from reads to peaks) in Galaxy. Peaks identified in chloroplast and mitochondrial genes were removed as the used ChIP protocol uses a nuclei isolation step and these binding events are considered background. For the PPD2 analysis, shared peaks were identified with at least 25% overlap using IntersectBed. Both ZmAN3 samples were normalised against the NLS-GS^{yellow} and NoAB controls. GO enrichments were analysed through the PLAZA platform 3.0 (Proost *et al.*, 2015), with a focus on biological process categories having log2 enrichments ≥ 1.4 and p value≤ 0.01. *De novo* motif analysis of DNA bound by PPD2 was carried out as previously described (Gonzalez *et al.*, 2015). Selection of the top 500 most enriched peaks was carried out as follows: first, all peaks per experiment were ranked according to their fold enrichment value, and the 500 highest values were marked. If a peak was marked in each experiment, it was considered a truely high enriched peak. Next, the selection per experiment was extended (to the 1006 highest

enriched peaks), until in total 500 peaks were found in common between all experiments among the highest enriched.

REFERENCES

Archacki R, Yatusevich R, Buszewicz D, *et al.* 2016. Arabidopsis SWI/SNF chromatin remodeling complex binds both promoters and terminators to regulate gene expression. Nucleic Acids Research, gkw1273.

Atanassov II, Atanassov II, Etchells J, Turner SR. 2009. A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods **5**, 14–11.

Binkert M, Crocco CD, Ekundayo B, Lau K, Raffelberg S, Tilbrook K, Yin R, Chappuis R, Schalch T, Ulm R. 2016. Revisiting chromatin binding of the Arabidopsis UV-B photoreceptor UVR8. BMC Plant Biology **16**, 103.

Buendía-Monreal M, Rentería-Canett I, Guerrero-Andrade O, Bravo-Alberto CE, Martínez-Castilla LP, García E, Vázquez-Ramos JM. 2011. The family of maize D-type cyclins: genomic organization, phylogeny and expression patterns. Physiologia Plantarum **143**, 297–308.

Bunnik EM, Le Roch KG. 2013. An Introduction to Functional Genomics and Systems Biology. Advances in Wound Care **2**, 490–498.

Cheeseman IM, Desai A. 2005. A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans. Science Signaling **2005**, pl1–pl1.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method forAgrobacterium-mediated transformation of*Arabidopsis* thaliana. Plant Journal **16**, 735–743.

Coussens G, Aesaert S, Verelst W, Demeulenaere M, De Buck S, Njuguna E, Inzé D, Van Lijsebettens M. 2012. Brachypodium distachyonpromoters as efficient building blocks for transgenic research in maize. J Exp Bot **63**, 4263–4273.

De Souza CP, Hashmi SB, Osmani AH, Osmani SA. 2014. Application of a new dual localization-affinity purification tag reveals novel aspects of protein kinase biology in Aspergillus nidulans. (J-H Yu, Ed.). PLoS ONE **9**, e90911.

Debernardi JM, Mecchia MA, Vercruyssen L, Smaczniak C, Kaufmann K, Inzé 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. The Plant Journal **79**, 413–426.

Dedecker M, Van Leene J, De Winne N, *et al.* 2016. Transferring an optimized TAP-toolbox for the isolation of protein complexes to a portfolio of rice tissues. Plant Molecular Biology **91**, 341–354.

Dedecker M, Van Leene J, De Jaeger G. 2015. Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry. Current Opinion in Plant Biology **24**, 1–9.

Dejonghe W, Kuenen S, Mylle E, Vasileva M, Keech O, Viotti C, Swerts J, Fendrych M, Ortiz-Morea FA, Mishev K, Delang S, Scholl S, Zarza X, Heilmann M, Kourelis J, Kasprowicz J, Nguyen le SL, Drozdzecki A, Van Houtte I, Szatmári AM, Majda M, Baisa G, Bednarek SY, Robert S, Audenaert D, Testerink C, Munnik T, Van Damme D, Heilmann I, Schumacher K, Winne J, Friml J, Verstreken P, Russinova E. 2016. Mitochondrial uncouplers inhibit clathrin-mediated endocytosis largely through cytoplasmic acidification. Nat Commun **7**, 11710.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. 2012. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research **41**, D808–D815.

Gadeyne A, Sánchez-Rodríguez 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.

Giepmans BNG, Adams SR, Ellisman MH, Tsien RY. 2006. The fluorescent toolbox for assessing protein location and function. Science **312**, 217–224.

Gonzalez N, Pauwels L, Baekelandt A, De Milde L, Van Leene J, Besbrugge N, Heyndrickx KS,

Cuéllar Pérez A, Durand AN, De Clercq R, Van De Slijke E, Vanden Bossche R, Eeckhout D, Gevaert K, Vandepoele K, De Jaeger G, Goossens A, Inzé D.2015. A repressor protein complex regulates leaf growth in *Arabidopsis*. The Plant Cell **27**, 2273–2287..

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. The Plant Journal **88**, 476–489.

Gregis V, Sessa A, Dorca-Fornell C, Kater MM. 2009. The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. Plant J. 2009 Nov;60(**4**):626-37.

Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY. 2001. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. Journal of Biological Chemistry **276**, 29188–29194.

Hoe Kim J, Tsukaya H. 2015. Regulation of plant growth and development by the GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR duo. J Exp Bot **66**, 6093–6107.

Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH. 2010. Identification and testing of superior reference genes for a starting pool of transcript normalization in *Arabidopsis*. Plant Cell Physiololgy **51**, 1694–1706.

Hsia AP, Chen HD, Ohtsu K, Schnable PS. 2010. DNA extraction from freeze-dried plant tissue with CTAB in a 96-well format. Cold Spring Harb Protocols **2010**, pdb.prot5516.

Karimi M, Bleys A, Vanderhaeghen R, Hilson P. 2007. Building blocks for plant gene assembly. Plant Physiology **145**, 1183–1191.

Kaufmann K, Muiño JM, Østerås 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.

Kim JH, Kende H. 2004. A transcriptional co-activator, AtGIF1, is involved in regulating leaf growth and morphology in *Arabidopsis*. Proceedings of the National Academy of Sciences **101**, 13374–13379.

Kobayashi T, Morone N, Kashiyama T, Oyamada H, Kurebayashi N, Murayama T. 2008. Engineering a novel multifunctional green fluorescent protein tag for a wide variety of protein research. PLoS ONE **3**, e3822.

Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vandekerckhove J. 1988. Determination of the Processing Sites of an *Arabidopsis* 2S Albumin and Characterization of the Complete Gene Family. Plant Physiology **87**, 859–866.

Lee DW, Lee S, Oh YJ, Hwang I. 2009. Multiple sequence motifs in the rubisco small subunit transit peptide independently contribute to Toc159-dependent import of proteins into chloroplasts. Plant Physiology **151**, 129–141.

Li S. 2015. The *Arabidopsis* thaliana TCP transcription factors: A broadening horizon beyond development. Plant Signaling & Behavior **10**, e1044192–00.

Lima M de F, Eloy NB, Pegoraro C, Sagit R, Rojas C, Bretz T, Vargas L, Elofsson A, de Oliveira AC, Hemerly AS, Ferreira PC. 2010. Genomic evolution and complexity of the Anaphase-promoting Complex (APC) in land plants. BMC Plant Biology **10**, 254.

Ma H, McLean JR, Chao LF-I, Mana-Capelli S, Paramasivam M, Hagstrom KA, Gould KL, McCollum D. 2012. A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms. Molecular & Cellular Proteomics **11**, 501–511.

Mohanty A, Luo A, DeBlasio S, Ling X, Yang Y, Tuthill DE, Williams KE, Hill D, Zadrozny T, Chan A, Sylvester AW, Jackson D. 2009. Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. Plant Physiology **149**, 601–605.

Morohashi K, Casas MI, Falcone Ferreyra ML, Mejía-Guerra MK, Pourcel L, Yilmaz A, Feller A, Carvalho B, Emiliani J, Rodriguez E, Pellegrinet S, McMullen M, Casati P, Grotewold E. 2012. A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. Plant Cell **24**, 2745–2764.

Murashige T, Skoog F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco

Tissue Cultures. Physiologia Plantarum **15**, 473–497.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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 BK, Cai X, Nebenfuhr A. 2007. A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. Plant Journam **51**, 1126–1136.

Palmer E, Freeman T. 2004. Investigation Into the use of C- and N-terminal GFP Fusion Proteins for Subcellular Localization Studies Using Reverse Transfection Microarrays. Comparative and Functional Genomics **5**, 342–353.

Paramban RI, Bugos RC, Wen Su W. 2004. Engineering green fluorescent protein as a dual functional tag. Biotechnology and Bioengineering **86**, 687–697.

Proost S, van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K. 2015. PLAZA 3.0: an access point for plant comparative genomics. **43**, D974–81.

Rhee SY, Mutwil M. 2014. Towards revealing the functions of all genes in plants. Trends in Plant Science **19**, 212–221.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology **17**, 1030–1032.

Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, Zhu JK, Ronald P, Fromm ME. 2006. Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. The Plant Journal **46**, 1–13.

Sarnowska E, Gratkowska DM, Sacharowski SP, Cwiek P, Tohge T, Fernie AR, Siedlecki JA, Koncz C, Sarnowski TJ. 2016. The Role of SWI/SNF Chromatin Remodeling Complexes in Hormone Crosstalk. Trends in Plant Science **21**, 594–608.

Schaeffer SM, Nakata PA. 2015. CRISPR/Cas9-mediated genome editing and gene replacement in plants: Transitioning from lab to field. Elsevier Ireland Ltd, 1–57.

Schäffer U, Schlosser A, Muller KM, Schafer A, Katava N, Baumeister R, Schulze E. 2010. SnAvi - a new tandem tag for high-affinity protein-complex purification. Nucleic Acids Research **38**, e91–e91.

Van Damme D, Coutuer S, De Rycke R, Bouget F-Y, Inzé 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, 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, Inzé 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. Nature Protocols **10**, 169–187.

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, Inzé D, Witters E, De Jaeger G. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis* thaliana. Molecular & Cellular Proteomics **6**, 1226–1238.

Van Leene J, Witters E, Inzé D, De Jaeger G. 2008. Boosting tandem affinity purification of plant protein complexes. Trends in Plant Science **13**, 517–520.

Varagona MJ, Schmidt RJ, Raikhel NV. 1992. Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein Opaque-2. The Plant Cell **4**, 1213–1227.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L,

Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. **26**, 210–229.

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, Inzé 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 Physiology **164**, 1122–1133.

Voss U, Larrieu A, Wells DM. 2013. From jellyfish to biosensors: the use of fluorescent proteins in plants. The International journal of developmental biology **57**, 525–533.

White DWR. 2006. PEAPODregulates lamina size and curvature in*Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **103**, 13238–13243.

White DWR. 2017. PEAPOD limits developmental plasticity in *Arabidopsis*. bioRxiv, 102707.

Wu Q, Luo A, Zadrozny T, Sylvester A, Jackson D. 2013. Fluorescent protein marker lines in maize: generation and applications. The International journal of developmental biology **57**, 535–543.

Yamaguchi N, Winter CM, Wu M-F, Kwon CS, William DA, Wagner D. 2014. PROTOCOL: Chromatin Immunoprecipitation from Arabidopsis Tissues. The Arabidopsis Book **12**, e0170.

Yun HY and Eipper BA. 1995. Addition of an Endoplasmic Reticulum Retention/Retrieval Signal Does Not Block Maturation of Enzymatically Active Peptidylglycine α-Amidating Monooxygenase. The Journal of Biological Chemistry **270**, 15412-15416.

SUPPLEMENTAL FIGURES AND TABLES

```
                                  20                                  40
                                   |                                   |
GFP  MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY 40
cYFP AVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY 40
                                  60                                  80
                                   |                                   |
GFP  GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK 80
cYFP GKLTLKFICT TGKLPVPWPT LVTTFGYGLM CFARYPDHMK 80
                                 100                                 120
                                   |                                   |
GFP  QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL 120
cYFP QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL 120
                                 140                                 160
                                   |                                   |
GFP  VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN 160
cYFP VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN 160
                                 180                                 200
                                   |                                   |
GFP  GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH 200
cYFP GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH 200
                                 220
                                   |
GFP  YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK 239
cYFP YLSYQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK 239
```

Figure S1. Allignement of GFP and cYFP amino acid sequences. The fluorescent molecules differ 6 amino acids from each other.

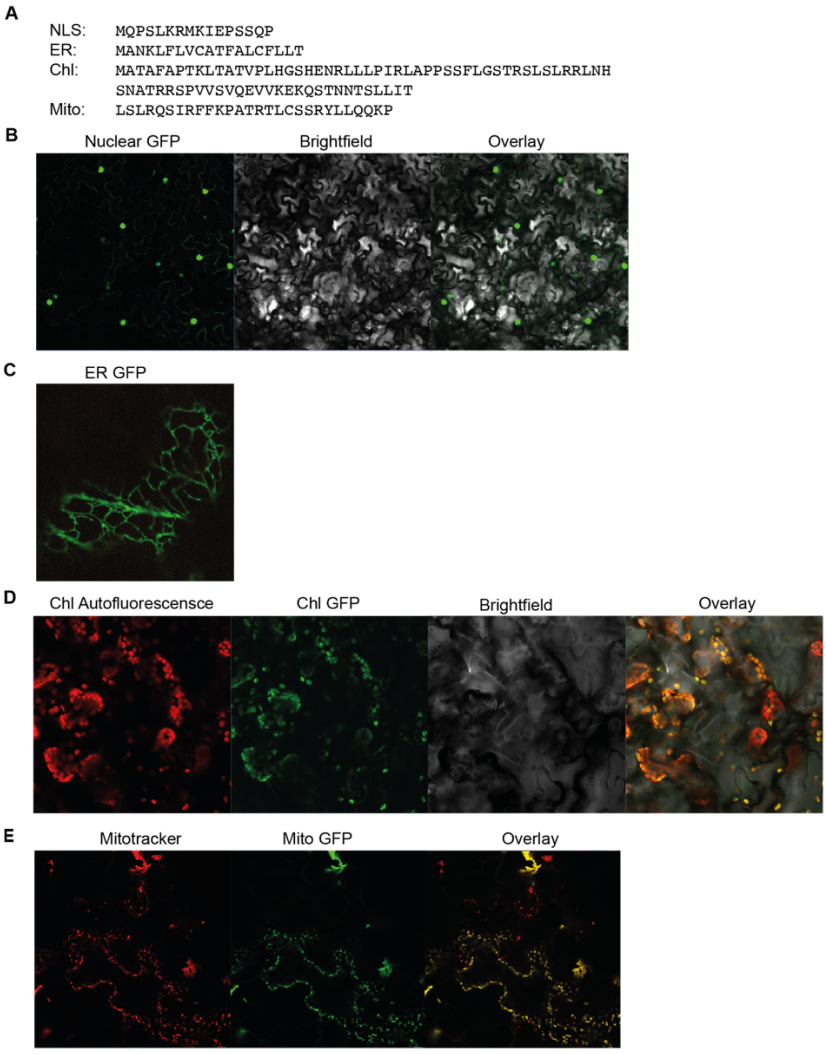


Figure S2. **Signal peptide TAP tag variants. (A-E)** Overview of the implemented signal peptide sequences and their respective localisation patterns. **(A)** Signal peptides were inserted N-terminally from the tag sequence. (**D)** Red signal originates from autofluorescence of the chloroplasts, panel four shows overlap of GS^{green} signal and autofluorescence signal. **(E)** Red signal originates from a mitotracker (see Materials and Methods). NLS: Nuclear Localisation Signal; ER: Endoplasmic Reticulum; Chl: Chloroplast; Mito: Mitochondria

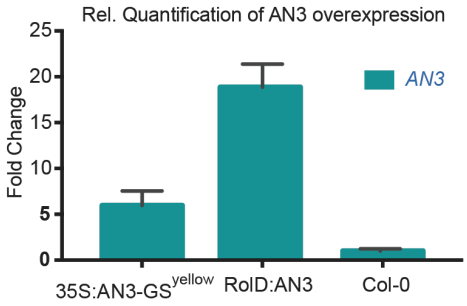


Figure S3. Relative quantification of AN3 overexpression by qRT-PCR in 6-days-old *Col-0* (WT) seedlings transformed with 35S:AN3-GS^{yellow} or RolD:AN3. Data normalised to the reference genes *PP2A* and *CDKA1;1*, and to the endogenous level of AN3 mRNA in WT. Error bars represent SE (n=3).

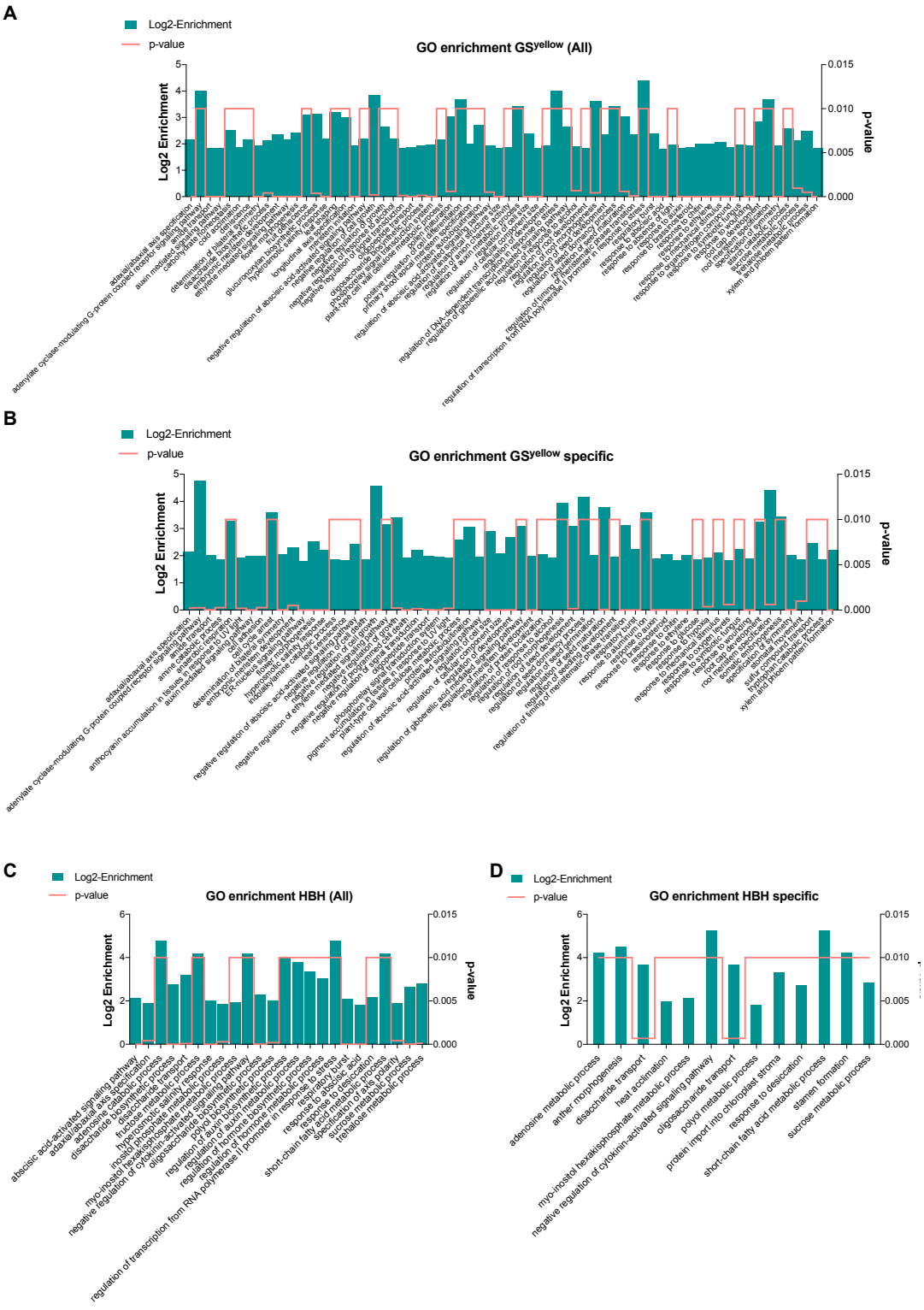

A
Log2-Enrichment
p-value
GO enrichment GS^yellow (All)
Log2 Enrichment
p-value
B
Log2-Enrichment
p-value
GO enrichment GS^yellow specific
Log2 Enrichment
p-value
C
Log2-Enrichment
p-value
GO enrichment HBH (All)
Log2 Enrichment
p-value
abscisic acid-activated signaling pathway
adaxial/abaxial axis specification
adenosine catabolic process
disaccharide biosynthetic process
fructose metabolic process
hyperosmotic salinity response
inositol phosphate metabolic process
myo-inositol hexakisphosphate metabolic process
negative regulation of cytokinin-activated signaling pathway
oligosaccharide biosynthetic process
regulation of polyol biosynthetic process
regulation of transcription from RNA polymerase II promoter in response to stress
response to abscisic acid
response to desiccation
short-chain fatty acid metabolic process
specification of cell polarity
sucrose metabolic process
trehalose metabolic process
D
Log2-Enrichment
p-value
GO enrichment HBH specific
Log2 Enrichment
p-value
adenosine metabolic process
anther morphogenesis
disaccharide transport
heat acclimation
myo-inositol hexakisphosphate metabolic process
negative regulation of cytokinin-activated signaling pathway
oligosaccharide transport
polyol metabolic process
protein import into chloroplast stroma
response to desiccation
short-chain fatty acid metabolic process
stamen formation
sucrose metabolic process

Figure S4. Histogram of the Gene Ontology (GO) enrichment of genes identified with ChIP-seq on cell cultures expressing PPD2-GS^{yellow} or PPD2-HBH as calculated through PLAZA v3.0, Dicots (Proost *et al.*, 2015). Only categories with a log2 enrichment >1.8 and p<0.01 are shown.

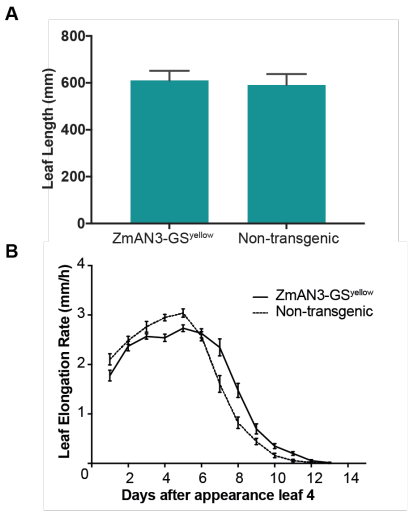


Figure S5. Phenotypical analysis of the transgenic ZmAN3-GS^{yellow} plants compared to their non-transgenic counterparts. **(A)** Final leaf length of leaf 4. **(B)** Leaf elongation rate of leaf 4. More than 15 plants were used per analysis. Error bars represent standard error.

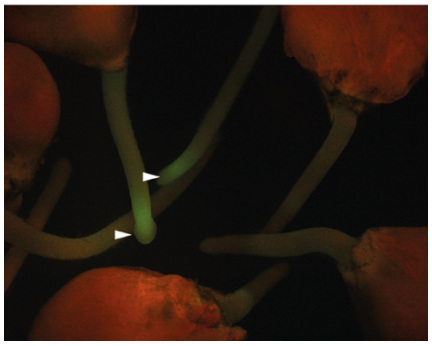


Figure S6. Screening of transgenic *Zea mays* root tips expressing pUBIL:ZmAN3-GS^{yellow} through fluorescence microscopy of the emerging root tips of *Zea mays*, 3 days after sowing. Arrowheads indicate pUBIL:ZmAN3-GS^{yellow} transgenic offspring as identified through the presence of the Yellow Fluorescent Protein signal.

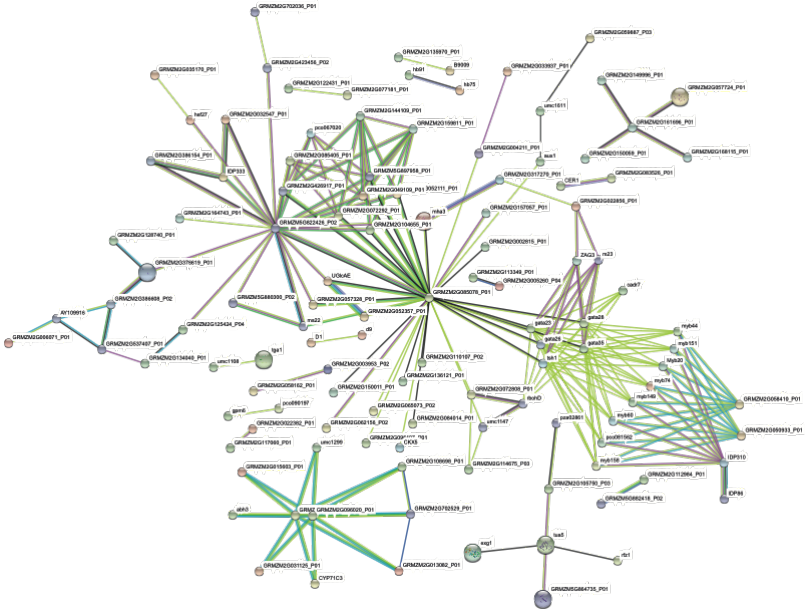


Figure S7. Visual representation of the connectivity of the 500 most enriched genes as identified by ChIP-seq on ZmAN3-GS^{yellow} in the growth zone of *Zea mays* leaf 4. Network constructed using the STRING database (http://string-db.org), with a medium confidence score (0.400) for associations. The nodes in the network show more connections (edges) than expected at random (p value 7,64e-12).

BED Annotation PLAZA
input BED file
output_summary
output_annotation
output_meta
FastQC:Read QC
Short read data from your current history
Contaminant list
html_file
MACS2
ChIP-Seq_Tag File
ChIP-Seq_Control File
output_bed_file
output_wiggle_file
output_narrowpeaks_file
output_broadpeaks_file
output_xls_b_interval_peaks_file
output_bdgpeng file
MarkDuplicates
Select SAM/BAM dataset or dataset collection
metrics_file
outFile
Filter SAM or BAM, output SAM or BAM
SAM or BAM file to Filter
Output alignments overlapping the regions in the BED FILE
output1
SortSam
Select SAM/BAM dataset or dataset collection
outFile
Map with BWA for Illumina
Forward FASTQ file
Reverse FASTQ file
output
Trimmomatic
Input FASTQ file (R1 first if pair)
Input FASTQ file (R2 second of pair)
fastq_out_r1_paired
fastq_out_r1_unpaired
fastq_out_r2_paired
fastq_out_r2_unpaired
fastq_out
Input dataset
output

Figure S8. Schematic representation of the ChIP-seq data processing pipeline in Galaxy**.** See Materials and Methods for specific settings used for peak calling.

Supplemental Table S1. TAP with ZmAN3 as bait. Additional proteins for which no orthologs were found in *Arabidopsis* TAP data. The data were compared to previously published results in *Zea mays* (Nelissen *et al.*, 2015).

		GS^{yellow}		Previously Reported	
Protein ID	**Protein Name**	**Division Zone(5)**	**Expansion Zone(5)**	**Division Zone(4)**	**Expansion Zone(4)**
GRMZM2G138419	Osmotin-like protein	5	5	4	4
GRMZM2G419806	oil yellow; chelatase subunit ChlI	1	4		
GRMZM2G009412	60S ribosomal protein L44	2	2		
GRMZM5G824600	Alcohol dehydrogenase class-3	2			
GRMZM2G112074	licensing factor MCM7	2	1		
GRMZM2G157061	factor isoform 4G	3			
GRMZM2G097511	Histone H2A	3	1		
GRMZM2G348666	Isoleucine--tRNA ligase	3	2		
GRMZM2G096123	Putative uncharacterised protein	3	2		
GRMZM2G144030	translation initiation factor 5A	3	3		
GRMZM2G152573	60S ribosomal protein L34	4	4		
GRMZM5G850966	60S ribosomal protein L36	4			
GRMZM2G389768	Cold shock	4	4		
GRMZM2G164020	Histone H1	4	5		
GRMZM2G063864	PHD finger protein ALFIN-LIKE	4	3		
GRMZM5G866100	Uridine kinase-like protein 2, chloroplastic	4	2		
GRMZM2G090422	60S ribosomal protein L13a	5	5		
GRMZM2G170336	40S ribosomal protein S20	5	4		
GRMZM2G059039	Apoptosis inhibitor 5	5	4		
GRMZM2G042008	ASPARTIC PROTEASE IN GUARD CELL 1	5	4		
GRMZM2G129007	ATP-dependent RNA helicase DHH!	5	3		
GRMZM2G010085	CHD3-type chromatin-remodeling factor PICKLE	5	4		
GRMZM2G145827	dynamin-related protein 1C-1E	5	5		
GRMZM2G094353	finger Ran-binding domain-containing protein 2	5	4		
GRMZM2G046932	Glutamine--tRNA ligase	5	4		
GRMZM2G084767	Importin subunit beta-1	5	5	4	3
GRMZM2G391364	Probable calcium-binding protein CML7	5	5		
GRMZM2G398527	Protein RRP5 homolog	5	5		
GRMZM2G461447	Putative E3 ubiquitin-protein ligase ORTHRUS 4	5	3		
GRMZM2G117900	Putative eukaryotic translation initiation factor 2 subunit 3-like protein	5	3		
GRMZM2G166345	translation elongation factor 1 epsilon-1	5	4		
GRMZM5G874163	Auxin Response Factor	(*)			

(*): Identified with one significant peptide, distinction between the different ARFs can not be made based on the detected sequence.

Supplemental Table S2. Growth zone-specific enrichment of interacting proteins of ZmAN3 in the leaf of *Zea mays*. Relative label-free quantification (MaxLFQ) of five independent TAP experiments on both dividing and expanding leaf tissue with the MaxQuant software. A paired *t*-test analysis was performed to identify significantly changing proteins between the division zone and the expansion zone.

Protein ID	Plaza Description	AT BHIF ID	Significant (*)	-LOG(P-value)(^)(y-axis in Figure 4B)	Difference (x-axis in Figure 4B)
GRMZM2G124566	GRF17/GRF18	AT2G36400	+	3.988	5.923
GRMZM2G034876	GRF1	AT3G13960	+	4.932	4.833
GRMZM2G178261	GRF15	AT2G22840	+	4.070	4.538
GRMZM2G067743	GRF11		+	5.757	4.403
GRMZM2G041223	GRF6	AT3G13960	+	3.027	3.691
GRMZM5G874163	ARF	AT2G33860	+	4.667	3.513
GRMZM2G105335	GRF3	AT3G13960	+	4.755	2.640

Significant (*): This column indicates which data points were significant in the two groups test.
-log10(p-value)(^): Test statistic, taking into account FDR=0.01 and S0=0.9
AT BHIF ID: Orthologue in *Arabidopsis* as determined through the PLAZA, Monocots v3.0 (Proost *et al.*, 2015) database using Best-hit Inparalog Families

Supplemental Table S3. List of primers used for cloning and qPCR

Cloning tags*	AttB2R_Gly4Ser_linker_Fwd	GGGGACAGCTTTCTTGTACAAAGTGGGTGGAG GCGGTTCAATGGACGAGAAGACCACC
	SBP+cYFP_Rev	TCCACCTCCACCTCCGGCTTCATCGTGTTGCCC AAGAGCCGCG
	SBP+cYFP_Fwd	CAACACGATGAAGCCGGAGGTGGAGGTGGAG CTGTGAG
	AttB3_GFP/cYFP_Rev	GGGGACAACTTTGTATAATAAAGTTGTCACTT GTACAGCTCGTCCATGCCG
	AttB4_omegaLeader_Fwd	GGGGACAACTTTGTATAGAAAAGTTGAACAAC AAACAACAAACAACATTCA
	AttB1R_NGSrhino/GS$^{green/yellow}$_Rev	GGGGACTGCTTTTTTGTACAAACTTGCTGAAC CGCCTCCACCCGG
Cloning ZmNLS	AttB1_ZmNLS_Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTAT GCCTACCGAGGAAAGAGTGAGG
	AttB2_ZmNLS_Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCA GTTCTTTCAGGTGAGCGGCTTTC
qPCR	AN3_Fwd	CAGGGAGAAGGAGGGTCACAC
	AN3_Rev	ATTGAAGATCGAGCCGCCATTAG
	GS^{yellow}_Fwd	GAGCCGCTTGAAGTTCTT
	GS^{yellow}_Rev	GGGAATAGCACCCTGAAA
	AN3_UTR_Fwd	GAAGCCGGAAATGGGAAGTG
	AN3_UTR_Rev	GCCACACTCTTTCAATTCCCA
	PP2AA3_Fwd	TAACGTGGCCAAAATGATGC
	PP2AA3_Rev	GTTCTCCACAACCGCTTGGT
	CDKA_Fwd	ATTGCGTATTGCCACTCTCATAGG
	CDKA_Rev	TCCTGACAGGGATACCGAATGC

* Due to the presence of the G4S linker in each tag universal primers can be used for cloning

Supplemental Table S4. MaxQuant search parameters

Group specific parameters	
General	
type	standard
multiplicity	1
variable modifications	Acetyl (Protein N-term), Oxidation (M)
digestion mode	specific
enzyme	Trypsin/P
max missed cleavages	2
instrument type	orbitrap
Label free quantification	
label free quantification	LFQ
LFQ min ratio count	1
fast LFQ	TRUE
LFQ min number of neighbors	3
LFQ av number of neighbors	6
First search	
separate var modifications first search	FALSE
separate enzyme first search	FALSE
Advanced	
main search ppm	4.5
max number of mod per peptide	5
max charge	7
individual pept mass tolerances	TRUE
Global parameters	
General	
	ZmB73_5a_WGS_translationsplus.fast
Fasta files	a
fixed modifications	Carbamidomethyl (C)
re-quantify	TRUE
match between runs	TRUE
matching time window	0.5
alignment time window	20
match unidentified features	FALSE
Sequences	
decoy mode	revert
special AAs	KR
include contaminants	TRUE
I=L	FALSE
Identification	
PSM FDR	0.01
Protein FDR	0.01
Site Decoy Fraction	0.01
min Pep Length	7
min peptides	1
min Razor+Unique peptides	1
min Unique peptides	1
min score unmodified peptides	0
min score modified peptides	40
min Delta score unmodified peptides	0
min Delta score modified peptides	17
base FDR calculations on Delta score	FALSE
filter labeled AA	TRUE
second peptides	TRUE
dependent peptides	FALSE
Protein Quantification	
min ratio count	1
peptides for quantification	Unique
use only unmodified peptides and	TRUE
modifications used in quantification	Acetyl (Protein N-term), Oxidation (M)
discard unmodified counterpart peptides	TRUE
Site Quantification	
mode	Use Least Modified Peptide
use for occupancies	Normalized Ratios

Label free quantification	
separate LFQ in parameter groups	FALSE
LFQ stabilise large ratios	TRUE
LFQ require MSMS	TRUE
Advanced	
calc peak properties	FALSE
mass difference search	FALSE
top x window	100
max peptide mass	4600
min time	NaN
max time	NaN
min peptide length for unspecific search	8
max peptide length for unspecific search	25
min peak length	2
cut peaks	TRUE
razor protein FDR	TRUE
MSMS	
Name	**FTMS**
MSMS Tolerance	20
Unit	Ppm
Top Peaks per 100 Da	12
De-isotoping	TRUE
HigherCharges	TRUE
Water Loss	TRUE
Ammonia Loss	TRUE
Dependent Losses	TRUE
Name	**ITMS**
MSMS Tolerance	0.5
Unit	Dalton
Top Peaks per 100 Da	10
De-isotoping	FALSE
HigherCharges	TRUE
Water Loss	TRUE
Ammonia Loss	TRUE
Dependent Losses	TRUE

Genetic manipulation avant la lettre

Simon Deraeck

INSERTION OF THE TAP TAG IN THE GENOMIC CONTEXT OF A GENE THROUGH TAC RECOMBINEERING

Nienke Besbrugge[1,2], Bernard Cannoot[1,2], Dominique Eeckhout[1,2], Nancy De Winne[1,2], Geert Persiau[1,2], Eveline Van De Slijke[1,2], Jelle Van Leene[1,2] and Geert De Jaeger[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2]Centre for Plant Systems Biology, VIB, 9052 Ghent, Belgium

Author contributions

G.D.J. and J.V.L. designed the research; N.B. performed most of the experiments and wrote the chapter; D.E., N.D.W., E.V.D.S. and G.P. carried out experiments and provided technical assistance; J.V.L, H.N and G.D.J. supervised and complemented the writing

ABSTRACT

Protein tagging in combination with recombinant protein expression is a powerful tool for functional analysis of proteins. This is reflected in the wide variety of tags and tag-based techniques that have been developed to unravel protein-protein and protein-DNA interactions, as well as to analyse protein localisation patterns. To be able to obtain single-cell resolution in such experiments it is important to clone the endogenous gene. Using overexpression promoters, for example, which are a common approach in recombinant protein expression, leads to loss of spatial and temporal information. However, selecting the correct promoter sequences for cloning can be challenging, as the endogenous regulatory landscape of plant genes is still poorly understood. The best way to overcome this problem is to tag the endogenous gene through homologous recombination strategies. Unfortunately, homologous recombination in plants occurs only with low efficiency, leading to the development of alternative cloning methods, such as the TAC recombineering system. This technique uses large transformation-competent artificial *Arabidopsis* chromosomes (TAC) in combination with an inducible homologous recombination system in *Escherichia coli* to generate gene tag fusions in a pseudo-chromosomal context. Here, we evaluate TAC recombineering to create fusion proteins with the GS^{yellow} TAP tag in *Arabidopsis*.

INTRODUCTION

FUNCTIONAL GENOMICS THROUGH TAGGING

Functional protein analysis through tagging requires the genetic insertion of the tag just next to the gene of interest. In general, two different strategies can be applied to produce such genetic fusions: classic molecular cloning and PCR-based methods. The first method uses the activity of certain restriction enzymes to cut the endogenous genomic DNA *in vitro* (Sharan *et al.*, 2009). This limits the technique through the necessity of restriction sites around the gene of interest. On the other hand, PCR based cloning techniques, including Gibson assembly as well as recombination based Multisite Gateway Vector^{TM} cloning technology (Invitrogen), also have their limitations. These mostly concern the length of fragments that are constructed, which cannot be too large and have to be combined through methods, such as adaptor-mediated extension overlap PCR (Atanassov *et al.*, 2009). Selecting the right sequences for cloning is also difficult: for example, when using only the coding sequence (CDS) of a gene, the endogenous regulatory sequences, found in introns and UTRs, necessary for transcriptional and post-transcriptional regulation are missing. It is therefore recommended to clone the fully predicted native gene instead of only the CDS.

Choosing the right promoter for functional protein analysis is even more challenging: a constitutive and strong promoter, such as the commonly used 35S promoter, causes a loss of spatial and temporal differences in expression profiles.

The protein under study will be expressed in places or developmental stages in which it is normally absent. Downstream analysis of interactions and localisation patterns is as such difficult, because novel false positive interactions are likely to be picked up. To circumvent this problem, endogenous promoters comprising all the native regulatory sequences can be used. However, despite the efforts of annotating the *Arabidopsis* genome, not much is known on the identity of plant promoters. It is often presumed that most promoter regulatory sequences lie in the 2-3 kb upstream of the start codon, or span until the stop codon of the upstream gene (which in *Arabidopsis* is indeed on average 1750 bp) (The Arabidopsis Genome Initiative, 2000). Using such a rather arbitrary cut-off demonstrates our lack of understanding in the epigenetic regulation of DNA transcription, as enhancer regions as well as terminator sequences are missing. It is therefore crucial to test the ability of the selected regulatory sequences to mimic the endogenous situation, which can be done in two different ways: either by comparing promoter:GUS expression data with, for instance, *in situ* mRNA hybridisation, or by expressing the recombinant protein in a knock-out mutant background. If the observed phenotype is complemented in the latter set-up, the protein copies its endogenous counterpart in both space and time. However, in case genetic mutants are not available (which is the case for most of the newly studied plant species, such as corn and other crops) or an obvious phenotype is missing for complementation analysis, this strategy cannot be applied. The situation gets even more complicated when studying large gene families with redundant functions, because they often need the generation of double or triple mutants. This is rather impossible if nothing is known about the function of the family. There have even been cases reported in which fusion proteins expressed from the same promoter show different expression patterns despite their ability to complement the observed phenotype (Zhou *et al.*, 2011*a*). This means that using genetic complementation as proof that the expression pattern represents the pattern of the endogenous gene can lead to mistakes.

As such, the best way to study gene functions is by genetically fusing the endogenous sequences of interest to the tag, which is possible by using a technique called homologous recombination. Here, a broad overview is given of this homologous recombination method for endogenous tagging, from its origin in yeast to its application in plants. Further, it is shown how the strategy can be applied to other plant proteins and tags, and possible pitfalls that arise during the procedure are discussed.

HOMOLOGOUS RECOMBINATION

The best way to preserve all endogenous regulatory sequences and reduce the number of steps for producing recombinant proteins is by tagging the gene of interest *in vivo.* This could be achieved by using the homologous recombination mechanism that repairs double-stranded DNA breaks (DSBR). In yeast, homologous recombination appeared to be very efficient for gene targeting (Szostak *et al.*, 1983), and it allowed the endogenous tagging of most protein-coding genes for a standardised functional analysis of the yeast proteome (Sarov *et*

al., 2006). Recombination to insert linear DNA is established in yeast through homologous DNA sequences as short as 50 bp or even smaller. However, in other model organisms and higher eukaryotes, homologous recombination appears to be less efficient, due to the need for longer homologous regions and a more complex integration mechanism (Zhou *et al.*, 2011*b*). One way to circumvent this problem is to construct artificial chromosomes consisting of long stretches of genomic DNA of the species of interest, and transform them into yeast. Such Yeast Artificial Chromosomes (YACs) have the advantage that they can contain megabase-sized genomic inserts, but the size comes at the cost of low concentrations and difficult purification (Copeland *et al.*, 2001). Improved YACs with enhanced copy numbers that contain only 80 to 700 kb have been successfully constructed for biolistic transformation of plants (Adam *et al.*, 1997). However, the clonal instability of YACs is rather high, as 40-50% of the subclones contain deletions due to the instability of long repetitive sequences. YACs are also often chimeric due to DNA coming from multiple genomic locations. As the constitutive, non-controllable activity of the yeast recombination machinery might enhance the mutation rate (Peterson *et al.*, 1998), the adaptation of YACs for functional analysis and transformation is rather slow (Zhou *et al.*, 2011*a*). Bacterial Artificial Chromosomes (BACs) and Phage 1 Artificial Chromosomes (PACs or fosmids), on the other hand, are generally smaller (with inserts of 200-300 kb or less), and can easily be produced in large quantities in *E. coli* (Copeland *et al.*, 2001; Sharan *et al.*, 2009).The availability of bacterial strains with an inactive (*i.e.,* RecA mutant) endogenous recombination machinery for BAC library maintenance, the higher stability of BACs compared to YACs and the lower mutation rates even after several rounds of replication have shifted the usage of YACs for engineering to BACs. Moreover, several BAC libraries for animals, plants and microbes, are already available.

Hence, the best way to tag genes endogenously is by using BACs in bacteria that have a homologous recombination mechanism. Although one could use the bacterial homologous recombination mechanism, *i.e.*, the RecA-dependent pathway, this system requires long stretches of homology (>500 bp) (Poteete and Volkert, 1988). Moreover, BAC containing strains are mostly $RecA^-$ to prevent rearrangements of DNA. Therefore, an inducible RecA system has to be inserted through a selectable plasmid that also contains the sequence of interest (Wang *et al.*, 2006). Phage specific recombinant systems, on the other hand, require less long homology arms (~50 bp or less), and provide the same recombination efficiency. Several phage systems have been explored, but the most successfully adopted one is the recombination method of the phage lambda Red (Zhang *et al.*, 1998; Muyrers, 1999).

LAMBDA RED RECOMBINEERING

Phage lambda has a well-studied recombination mechanism that uses three proteins for homologous recombination of linear dsDNA: Gam, Exo and Beta (=PL operon). The Gam protein helps to inhibit the bacterial-specific defence mechanism against the insertion of linear DNA carried out through the RecBCD proteins. This

allows the electroporation of linear dsDNA (for instance, an amplified PCR product), which is then processed by the Exo protein that has a 5′- 3′ exonuclease activity and will produce 3′ linear ssDNA overhang ends that are recognised and bound by the Beta proteins for protection. Binding of Beta will also enhance the annealing of the complementary single strands, leading to strand exchange necessary for recombination (Sawitzke *et al.*, 2007; Mosberg *et al.*, 2010). To circumvent the need of co-transformation of a plasmid that contains these three genes under the regulation of an inducible promoter (Zhang *et al.*, 1998; Muyrers, 1999), a specific *E. coli* strain has been made. This strain, DY380, is a derivative of the DH10B strain that is typically used for BAC maintenance. It has the recombination cassette of a defective lambda prophage inserted into the genome. This cassette is placed under the regulation of a Lambda PL promoter bound by a temperature-sensitive lambda cI857 repressor (Yu *et al.*, 2000). Adaptation of this new strain for endogenous tagging in a chromosome-like background has significantly helped increase our understanding of proteomes in a system-wide manner, and many applications, such as the GFP tag-based unraveling of the *C. elegans* proteome (Sarov *et al.*, 2006), have been reported. Besides tagging, recombineering also allows easy and precise manipulation of genomes, which was shown to be very efficient in mice (Copeland *et al.*, 2001; Liu, 2003; Swing and Sharan, 2004) where it helped to construct a library of knock-out mutants that can be used in functional genomics (Bradley *et al.*, 2012).

RECOMBINEERING IN PLANTS

The availability of the *Arabidopsis* Col-0 genome in Transformation Competent Artificial Chromosomes (TAC) as a JAtY library (constructed at the John Innes Centre), has enabled the adaptation of the Lambda Red-based recombineering technique for plants. The TAC library (36864 clones in total) has an average insertion size of 68 kb, and the corresponding sequences, which cover >90% of all *Arabidopsis* genes, are very well documented (Hirose *et al.*, 2015). Because of the nature of the used backbone vector, the pYLTAC17, the JAtY library can easily be used for complementation experiments in T-DNA mutant lines. pYLTAC17 is also a binary vector, containing two *ori*'s for propagation in both *E. coli* and *Agrobacterium tumefaciens* to allow *Agrobacterium*-mediated transformation. However, some problems have been reported regarding the transformation efficiency of the JAtY library into bacteria (Zhou *et al.*, 2011*b*).

Implementation of homologous recombination for the production of whole-gene reporter fusions has been systematically evaluated in *Arabidopsis* with the use of SW102 *E. coli* cells and galK selection. The outline of the original protocol is depicted in Figure 1: first, the TAC that contains the gene of interest approximately in the middle has to be selected from the JAtY library. This is easily achieved through a web tool (www.arabidopsislocalizome.org) (step 1). The central position of the gene in the TAC is important to make sure that the DNA can easily be accessed for strand invasion to enhance the efficiency of homologous recombination. It also ensures that the gene resides in its endogenous context. The

TAC clone is then extracted from the DH10 *E. coli* strain, and transferred to the recombineering-competent SW102 strain (step 2). These SW102 cells contain a complete galactose operon, except for the galK gene in addition to the recombineering machinery. At the same time, the linear dsDNA cassette that carries the galK selection marker has to be amplified with primers containing 50-bp long arms homologous to the sequences flanking the desired insertion site. To circumvent the need for tag cassette-specific primers, a short linker was added to the cassette, which also provides a flexible poly-glycine and poly-alanine arm between the bait and the tag at the same time (Zhou *et al.*, 2011*a*). The resulting dsDNA fragment is then transferred to the TAC-containing strain by means of electroporation, and the lambda Red recombination machinery is switched on via a heat shock, leading to homologous recombination (step 3). By plating out the bacteria on minimal medium with galactose, the resulting recombinants can be selected through the galK marker present in the cassette that will carry out the first step in the conversion of galactose to glucose. Hereafter, a second round of recombineering follows, but this time an amplified GFP-containing cassette is transferred into the bacteria (step 4). Because the galK selection marker is now lost, negative selection has to be carried out using medium that contains 2-deoxy-galactose with glycerol. Finally, the GFP-tagged TAC is extracted and transferred into *Agrobacterium* (step 5) to allow plant transformation through floral dip (Clough and Bent, 2008) (step 6).

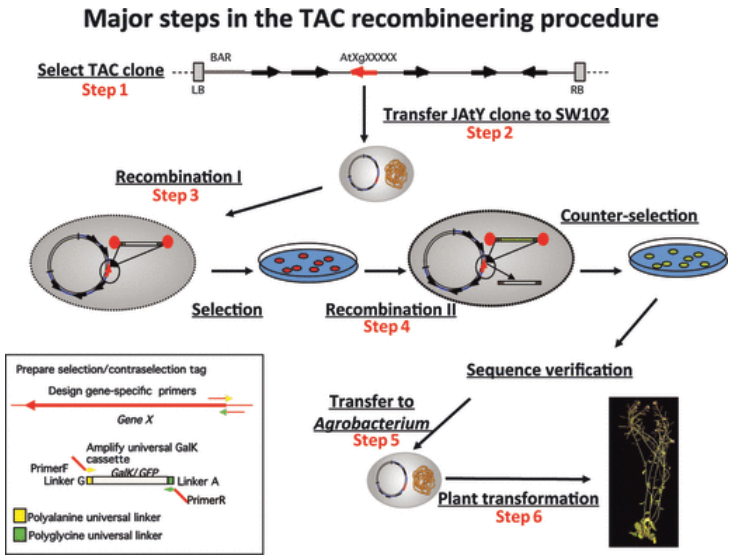


Figure 1. Schematic overview of the original TAC recombineering procedure for plants. All steps in the procedure are depicted. Left panel shows the strategy for primer design. Yellow and green boxes represent linker sequences; red lines indicate gene-specific primers. Figure taken from (Zhou *et al.*, 2011*a*).

Based on the reported success of GFP tagging through homologous recombineering in *Arabidopsis* BAC clones (Zhou *et al.*, 2011*a*), we wanted to see whether the protocol can be adapted to other tags, *i.e.*, our TAP tags. To this end,

the TOR (Target Of Rapamycin) protein was selected as bait. TOR is a kinase protein that has been heavily studied in animals because of its function in sensing the nutritious state of the cellular environment, and the linked regulation of cell growth, proliferation and survival.

In animals, mTOR was found to be part of two different complexes (mTORC1 and mTORC2), consisting of mTOR1, RAPTOR (Regulatory-associated protein of TOR) and mLST8 (Lethal with Sec 13) or mTOR2, RICTOR (Rapamycin-insensitive companion of TOR) and mLST8, respectively(Loewith *et al.,* 2002; Wullschleger *et al.,* 2005). TORC1 is bound by FKBP12 (FK506-binding protein) specifically upon rapamycin treatment (Loewith *et al.,* 2002). This inhibits TOR activity, causing a stop in cell cycle progression (Loewith *et al.,* 2002; Beauchamp and Platanias, 2012). Both TOR complexes are involved in several cellular processes, such as protein translation, ribosome biogenesis, actin dynamics and autophagy (Wullschleger *et al.,* 2006; Nietzsche *et al.,* 2016). Although the TOR kinases in eukaryotic species seem to be highly conserved, so far only one TOR complex was found to be present in plants (John *et al.,* 2014). Its central function in growth regulation in relation to nutrient levels is preserved in plants, and homologues of the known interactors RAPTOR, LST8 and one of the most prominent substrates of the TOR complex, ribosomal S6 kinase (S6K), are present in *Arabidopsis* (Deprost *et al.,* 2007). Previous work in *Nicotiana benthamiana* further confirmed the physical interactions between TOR, RAPTOR and S6K in plants via transient overexpression, and showed that these interactions lead to phosphorylation of the S6K (Rexin *et al.*, 2015). Phosphorylated S6K in turn can target RBR, which relocates to the nucleus upon activation. There, RBR associates with E2FB, inhibiting binding of E2FB to the promoter of genes involved in the cell-cycle (Henriques *et al.,* 2014). This provides a direct link between TOR and cell growth. Another cytosolic target of S6K is Ribosomal Protein S6 (RPS6), a component of the 40 S ribosomal subunit. Activation of RPS6 by S6K enhances translation of rDNA genes, providing a direct link between TOR, stress signals and the translational and transcriptional regulation of plant growth.

Previous in house TAP purifications on an *Arabidopsis* cell suspension culture overexpressing TOR-GS^{rhino} could not confirm any of the predicted interactors. The very long TOR-coding sequence gave a lot of cloning problems, and we could not exclude that the implemented CDS was incorrect. Therefore, we decided to insert the tag into a TOR-containing TAC clone to retry the TOR complex purification and characterisation.

In parallel, the same approach was used for the endogenous tagging of four other bait proteins (AN3, APC10, AURORA1 (AUR1) and TPLATE) with the new multifunctional GS^{yellow} tag (see previous chapter). This chapter reports on the encountered problems, the tested trials for solution, and the conclusions for TAC-based recombineering strategies in functional genomics experiments in plants.

RESULTS

ADAPTATION OF THE RECOMBINEERING PROTOCOL FOR TAP-TAGGING IN *ARABIDOPSIS*

We wanted to assess whether the recombineering technique could also be used for the addition of a TAP tag in a chromosome-like environment. To anticipate the reported inefficiency of the *GFP* recombineering step, a different *E. coli* strain was selected (SW105), which harbours an inducible *flippase* gene inserted into its genome. A cassette was then constructed that consists of a TAP tag (GS^{rhino} or GS^{yellow}) and an ampicillin resistance marker (AmpR) flanked by two flippase recognition targets (FTR) instead of the original *galK* selection system. A visual representation of the whole recombineering procedure can be found in Figure 2.

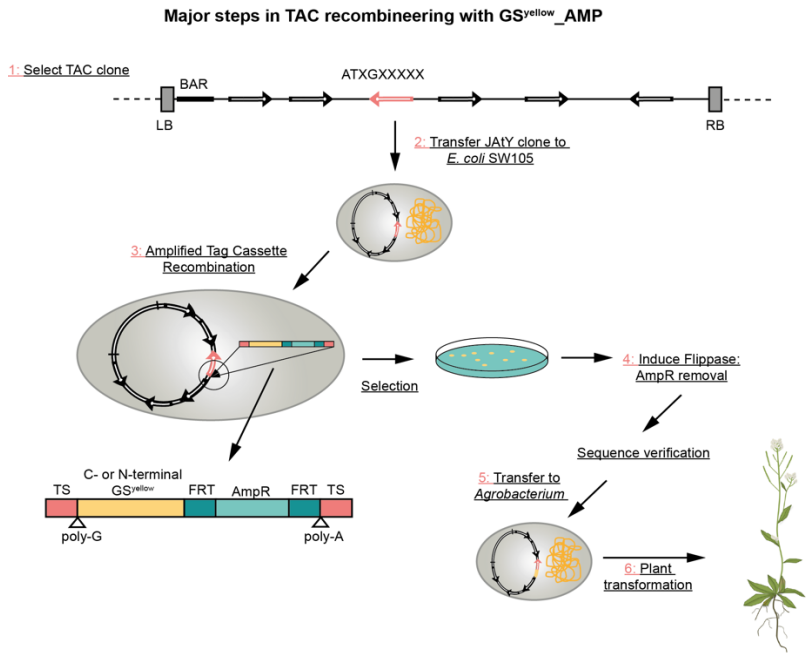


Figure 2. Adapted TAC recombineering scheme for TAP tagging in plants. The two original recombineering events were reduced to one, and selection was carried out using an ampicillin resistance marker (AmpR). Removal of this marker is done via the flippase recognition targets (FRTs), before transfer of the clone to *Agrobacterium*. The GS^{yellow}_AMP recombineering cassette is depicted on the left.

To validate this new recombineering strategy, the TOR protein was tagged at the N terminus with GS^{rhino} in the TAC clone JAtY77P22 selected from www.arabidopsislocalizome.org, and expressed in a WT Col-0 background. All steps in the recombineering approach were successful using standard settings, and 18 T1 plants could be generated. From these 18 plants, six were found to have a single locus insertion, and two independent lines with accumulation of the recombinant

protein on Western blots (lines 4.2 and 5.1, see Figure S1) were selected. Subsequent duplicate TAP purifications on these lines delivered the known interactors LST8-1 and RAPTOR1B, as well as an interactor that is reported to bind to the mTOR complex in animals (see Table 1). The latter protein, which is an orthologue of the mammalian Tti1 (Telo2 Interacting Protein1), has a stabilising effect on members of the phosphatidyl inositol 3-kinase-related protein kinase (PIKK) family, to which TOR belongs {Kaizuka:2010cd}. This stabilising effect occurs in cooperation with telomere maintenance2 (Tel2), another interactor of Tti1 (Takai *et al.*, 2010). Tel2 preferentially binds to freshly translated PIKK member proteins, helping maturation of the complexes they belong to, such as the mTORC complexes. Tti1 and its homologue Tti2 form a complex with Tel2 that is called the TTT complex {Kim:2013je}. Binding of this complex to RUVBL1/2, an ATPase, provides a sensor for energy stress. Energy stress will release RUVBL1/2 from the complex and result in loss of the interaction between Tel2 and mTOR {Kim:2013je}. In maize, LOF mutants for ZmTti2 were found to have smaller seeds due to lower amounts of endosperm per seed. This was shown to be due to a lack in detectable levels of ZmTOR, suggesting that stabilisation of PIKK complexes during protein folding through the TTT complex is conserved between animals and plants {Garcia:2017ki}.

These results thus show the successful adaptation of the TAC recombineering method for GS^{rhino} TAP tagging in *Arabidopsis.* Recombineering the tag into the TAC clone and transforming it into *Arabidopsis* appeared to be quite straightforward. Moreover, known interaction partners of TOR could be identified when TAP was performed in *Arabidopsis* seedlings.

Table 1. Interaction partners of TOR identified through TAP experiments on *Arabidopsis* seedlings

TOR		Seedlings	Previously reported
ProteinID	**Protein name**		
AT1G50030	TOR, target of rapamycin	2	x
AT3G08850	RAPTOR1B	2	x
AT3G18140	LST8	1	x
AT1G79190	Tti orthologue Tti-1	1	x

The list contains all the identified interaction partners from TAP on two biological repeats transformed with TAC TOR-GS^{rhino}

RECOMBINEERING WITH GS^{YELLOW}

Next, the same TAC recombineering set-up was applied for the tagging of AN3, APC10, AUR1 and TPLATE in their respective JAtY clones (see Figure S2). A GS^{yellow}_AMP cassette was first generated consisting of the GS^{yellow} tag and the ampicillin selection marker (AMP). This GS^{yellow}_AMP cassette was then used to add 50-bp homologous arms for C-terminal (AN3, AUR1 and TPLATE) or N-terminal tagging (APC10) (hereafter referred to as GS^{yellow}_AMP (50)). Despite several attempts, the GS^{yellow}_AMP (50) recombineering method only worked for APC10 TAC and AN3 TAC.

For all four constructs, the isolation of the TAC clone from the DB10 *E. coli* strain was successful, but the transfer to the competent SW105 cells was difficult for the AUR1 TAC clone (*i.e.*, JAtY80G11). Several attempts resulted in bacterial colonies with two distinct phenotypes when selecting for AUR1 TAC. To eliminate the possibility of a contaminated starter DB10 strain containing two different AUR1 TACs, a new clone was prepared. However, producing SW105 cells containing the necessary TAC clones appeared to be very difficult and required a lot of hands-on work, and the amount of obtained colonies varied substantially. Another issue was the loss of the TACs during propagation in SW105 cells, seen upon testing the colonies a second time for the presence of the TAC after restreaking on fresh selection medium.

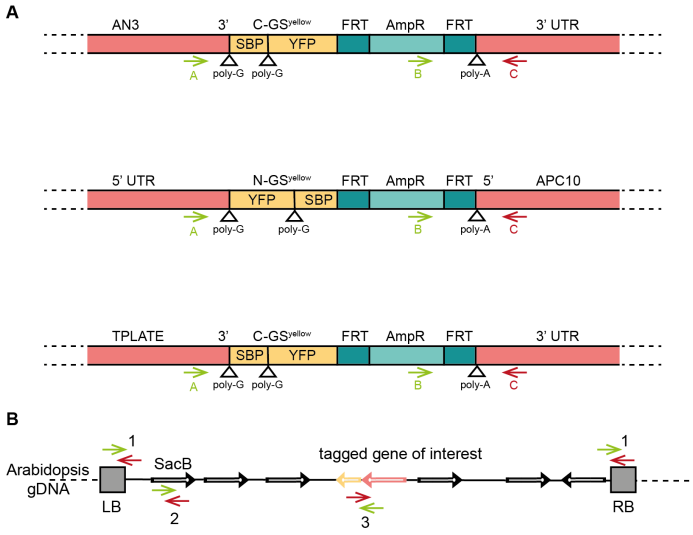


Figure 3. Overview of sequencing strategies and primer positions. (A) Analysis of GS^{yellow}_AMP insertion through recombineering for TAC AN3, APC10 and TPLATE. **(B)** Validation of full insertion of the tagged TAC clone into the *Arabidopsis* genome after *Agrobacterium*-mediated transformation. 1: border-specific primers; 2: *SacB* specific primers and 3: gene tag-specific primers. Primer sequences are provided in Table S1. FRT: Flippase Recognition Target; SBP: Streptavidin-Binding Peptide; YFP: Yellow Fluorescent Protein; AmpR: Ampicillin Resistance Marker.

Finally, only for APC10 TAC and AN3 TAC recombineering the GS^{yellow}_AMP (50) into the TAC resulted in viable colonies on the ampicillin selection medium (84/88 for APC10, and 34/96 tested colonies for AN3). Sequencing the inserted cassettes with construct specific test primers A, B and C (see Figure 3 and Table S1) showed insertion of the GS^{yellow} TAP tag in the correct reading frame with no nucleotide sequence errors (Figure S3). Because our GS^{yellow}_AMP cassette is bigger (2040 bp) than the original *galK* cassette (1231 bp) and 50 bp homologous ends only work well with insert sizes up to ~1000 bp (Zhou *et al.*, 2011*a*), the GS^{yellow}_AMP(50) cassettes for TPLATE and AUR1 were re-amplified using primers that extended the homologous arms to 200 bp. Indeed, this helped the recombineering for the TAC TPLATE (41/96 positive colonies), but when the DNA of these colonies were

sequenced, it became clear that the SBP part of the GS^{yellow}_AMP sequence was missing (Figure S4**)**. More colonies were therefore prepared. Here, sequencing revealed correct insertion of the cassette at the C terminus when using primer B, but not with primer C (see Figure S4). Due to the limited amount of time and the fact that this problem seemed to be present in every tested clone, the recombineering of GS^{yellow}_AMP in TAC TPLATE was stopped at this point. For the TAC AUR1, all attempts to insert the tag cassette failed, so also this construct was not continued. For TAC AN3 GS^{yellow}_AMP and TAC APC10 NGS^{yellow}_AMP, inducing the flippase gene through arabinose treatment, which causes loss of the AmpR gene, worked efficiently. Correct removal of *AmpR* was again confirmed through sequencing (Figure S5 shows the result for TAC AN3 GS^{yellow}). Transfer of TAC AN3 GS^{yellow} to *Agrobacterium* was successful for 71/96 of the tested colonies, and for TAC AP10 GS^{yellow}, all tested colonies were positive (18 in total).

In conclusion, despite the optimisation of both the experimental transformation procedure as well as the production of the recombineering cassette, only the TAC APC10 NGS^{yellow} and TAC AN3 GS^{yellow} constructs could be successfully produced. For the TAC AUR1 construct, correct recombineering failed. Taking all these results together, it is clear that the success rate varies for different TAC clones concerning the transfer of the vector into the SW105 strain and upon recombineering of the tag cassette.

TRANSFORMATION OF THE TAGGED TAC CLONES INTO *ARABIDOPSIS THALIANA*

Retaining two successfully tagged vectors, we transformed both constructs into their respective mutant backgrounds (*i.e.*, *an3-4* and *apc10-2*$^{+/-}$) (Horiguchi *et al.*, 2005; Eloy *et al.*, 2011). Here, T1 transformants were selected. The initial selection on Basta containing growth medium (through the *BAR* gene) provided resistant plants that did not appear to have taken up the TAC, as analysed through PCR. Strangely, repeating the PCR on these plants with Basta-specific primers showed a fragment of the correct size for some T1 transformants, indicating that the Basta marker on the TAC was indeed present but that the remainder of the TAC was deleted (data not shown). Therefore, a second method was applied in which the seeds are directly grown in soil and the plants sprayed with Basta (Logemann *et al.*, 2006). Again, this resulted in false positive plants (three for both constructs). Finally, an adapted selection method was used as published by Dr. Alonso's group who developed the TAC library (see Materials and Methods) (Zhou *et al.*, 2011*a*). In total, three and two T1 transformants could be selected that showed genomic insertion of TAC APC10 NGS^{yellow} and TAC AN3 GS^{yellow}, respectively. Next, the full insertion of the TACs was checked through PCR on genomic plant DNA with (i) border-specific primers that amplify the insertion of the TAC into the pYLTAC17 vector (left and right), (ii) *SacB* (adjacent to the left border of the TAC) specific primers and (iii3) gene tag specific primers (see Figure 3 and Table S1). Only one of the two TAC AN3 GS^{yellow} T1s contained the *BAR* gene (confirmed through PCR), but both T1s failed to show the presence of any of the anticipated PCR fragments. For the TAC APC10

NGS^{yellow}, all three T1s contained the tagged APC10-NGS^{yellow} and the left border of the T-DNA. However, only one T1 also contained the *SacB* gene and the right border, whereas the other two T1s clearly showed major deletions of the TACs despite positive selection through the *BAR* gene. No new attempts were carried out for the AN3 construct. Moreover, the transformation of the recombinant constructs into plants appeared to be very difficult. These results and the limited time frame for the construction of the recombinant vectors led to the preliminary termination of all further TAC-based cloning efforts.

COMPLEMENTATION ANALYSIS OF *apc10-2* TRANSFORMED WITH TAC APC10 NGS^{YELLOW}

The *apc10-2* mutant was previously reported to show female gametophyte lethality, *i.e.*, aberrant ovule formation (Eloy *et al.*, 2011). As such, self-fertilisation of the hemizygous insertion mutant always results in a 1:1 segregation ratio of wild-type offspring versus offspring of plants hemizygous for the *apc10-2* T-DNA insert. To see whether the TAC APC10 NGS^{yellow} construct could complement the *apc10-2* phenotype, we first had to make sure that the T1 transformant was hemizygous for the T-DNA insert in the *APC10* gene. This appeared not to be the case, which suggest that a WT plant was used for floral dip instead of a hemizygous *apc10-2* mutant. Therefore, the TAC APC10 NGS^{yellow} T1 was pollinated with pollen from a hemizygous *apc10-2* plant. The offspring (T2) was selected on medium containing both the selection for the tag (Basta, 20 µg/mL L-phosphinothricin) as for the SALK T-DNA insertion (kanamycin, 50 µg/mL), and was self-fertilised to produce seeds. Unfortunately, the segregation ratio of the T3 seedlings on kanamycin stayed at 1:1, which was further confirmed by genotyping PCR with the *apc10-2* SALK primers. No homozygous mutant T3 plants were found, indicating that the inserted TAC APC10 NGS^{yellow} construct could not complement the aberrant ovule formation. Given the low success rate for the AN3, AUR1 and TPLATE constructs and the lack of complementation for TAC APC10 NGS^{yellow} in the T3 plants, no further effort was put into unravelling at what level the tagged APC10 was not able to complement (*i.e.*, mRNA versus protein level).

DISCUSSION

AN ADAPTED SELECTION SCHEME FOR THE RECOMBINEERING OF TACS WITH TAP TAGS: A TAC TOR GS^{RHINO} CASE STUDY

To circumvent the difficulties reported with recombineering when using the *galK* positive/negative selection system and two subsequent recombineering events, as well as to provide a faster, more high-throughput approach (Warming, 2005), we slightly adjusted the published recombineering method with TACs for plants (Poser *et al.*, 2008; Zhou *et al.*, 2011*a*). Through the successful TAC TOR NGS^{rhino} construct, we showed here that the insertion of a selectable marker into the tag cassette indeed allowed easy screening of TAP-tagged recombineered *E. coli* colonies (Poser *et al.*, 2008). Moreover, losing the selection marker before transfer to *Agrobacterium* turned out to be very efficient, thanks to the insertion of

two recognition sites adjacent to each side of the marker in combination with the *flippase*-containing SW105 strain (Lee *et al.*, 2001; Warming, 2005; Poser *et al.*, 2008; Sharan *et al.*, 2009).

The functionality of the tagged protein was evaluated through the identification of its interaction partners using TAP coupled to LC-MS/MS. After removal of the contaminants, a list of specific interactors found per tag is obtained for each bait protein. By picking up the known interactors RAPTOR1B (Mahfouz, 2006) and LST8-1 (Moreau *et al.*, 2012), we confirmed the correct production of the recombinant TOR-GS^{rhino} protein (see Supplemental Data S1). However, because the protein was expressed in a WT Col-0 background, it is not possible to conclude that the fusion protein is fully functional or that the DNA in the TAC contains all the necessary remote regulating *cis*-elements. Besides complementation analysis of a TOR mutant plant, one could also test the activity of the tagged TOR by analysing its capacity to phosphorylate the S6K1 kinase (Xiong and Sheen, 2012), a known target of the TORC that links TOR to the regulation of elongation of transcription. However, making a distinction between endogenous versus tagged TOR activity *in vivo* would be rather difficult, and was therefore not tested. Nevertheless, the reported stabilising interaction of the Tti1 orthologous protein TTI, which was found to interact with the TORC in animals (Kaizuka *et al.*, 2010) and maize (Garcia *et al.*, 2017), is also seen here. However, the fact that the interaction with LST8-1, a core member of the TOR complex, was only found in one of the two experiments might indicate that the endogenous promoter activity is too low, and that competition of the tagged protein with the endogenous counterpart for complex formation interferes with detection of interaction partners. However, it has also been suggested that analysing the TOR complex through purification might be easier when pulling down the complex through a tagged LST8 protein (Sarbassov *et al.*, 2011), as purifying complexes with high molecular weight proteins, such as TOR, is relatively difficult (Sarbassov *et al.*, 2011), own unpublished data).

Overall, as the recombineering of TAC TOR with the GS^{rhino} TAP tag was only used to confirm the usability of our adapted recombineering approach and some known interactors were picked up with TAP in seedlings, we conclude that the approach could work and can be expanded to other TAP tags and proteins.

RECOMBINEERING WITH THE GS^{YELLOW} TAP TAG

Despite the relatively easy insertion of the GS^{rhino} TAP tag into the TAC of TOR with the adapted protocol, some problems were encountered when using recombineering for tagging with the GS^{yellow} TAP tag cassette. Difficulties occurred mostly when transferring the TAC clone from the DH10B to the SW105 strain (TAC AUR1), and recombineering repeatedly failed to produce positive GS^{yellow}_AMP containing colonies. Although the transfer of large DNA to SW105 is supposed to be successful in 98% of all cases, problems with confirming the JAtY presence due to the nature of the used JAtY clone have been previously reported (Zhou *et al.*, 2011*b*). Reordering a fresh clone (which eliminates the chance of loss of the JAtY clone over time) from the JAtY library helped here to proceed to the

recombineering step, but we cannot exclude that simply selecting another clone for recombineering could also have solved the problems with DNA transfer (Zhou *et al.*, 2011*b*). Regarding the problems with the recombineering, low quality of the amplified GS^{yellow}_AMP cassette or the prepared electro-competent cells could here be the reason why the correct insertion failed (Poser *et al.*, 2008; Zhou *et al.*, 2011*a*). This could be improved in the future by amplifying the cassette multiple times, and combining and enriching the cassette DNA to obtain a higher final concentration for recombineering. Additionally, a second purification round to purify the PCR product could also increase the quality, although one has to keep in mind that this will also lead to some loss of the PCR product, therefore requiring an even higher start concentration. For the TAC TPLATE construct, using longer homologous arms (up to 200 bp) significantly increased the recombineering efficiency.

CORRECT INSERTION OF THE TAG CASSETTE AND TRANSFER OF THE RECOMBINANT VECTOR INTO *AGROBACTERIUM*

Verifying the correct insertion of the tag cassette after recombineering is a very crucial step in the protocol (Poser *et al.*, 2008). Although no mutations were found to be present in the cassettes as such, partial loss of the tag sequence, namely the SBP part, was seen for the first TAC TPLATE GS^{yellow}_AMP recombineering attempt. This might be due to the presence of the same polyG linker both at the end of the cassette and in the middle of the GS^{yellow} tag (between the SBP and the YFP gene), an easy problem to solve by designing primers to pick up the GS^{yellow}_AMP cassette with longer homology that stretch along the 5′ sequence (the SBP part) of the tag cassette. Indeed, the second attempt showed that this problem as such was resolved, but now the occurrence of two versions of the GOI (one tagged versus one untagged) in a single bacterial colony interfered with the correct production of the TAC TPLATE GS^{yellow} vector. It was however not clear if this duplicity was due to the primers used for sequencing, or to the actual presence of two TACs or a rearrangement of some TAC DNA in the bacterial cell. As TACs have an average insertion size of 63 kb, sequencing the whole vector would be required to unravel the nature of this erroneous result. Given that this type of problem was not found in the two other constructs (AN3 and APC10) and that BAC based systems are reported to be single colony - single vector-like (Copeland *et al.*, 2001), it might not reflect common errors encountered with TAC recombineering. Subsequent transfer of the correct vectors for AN3 and APC10 appeared to be more efficient than the expected 73% success rate (Zhou *et al.*, 2011*b*), although, using only two constructs, we have to be careful with generalising these observations.

ARABIDOPSIS TRANSFORMATION AND COMPLEMENTATION

In the end, two out of four of the intended constructs could be made using the recombineering method, and one of these constructs showed major deletions when transferred to plants. As this is a known characteristic of working with such large pieces of DNA in TAC clones, it is advisable to screen at least 10 to 20 T_1 transformants to make sure that about five single-locus lines will be retained. Given

that the TAC AN3 is smaller in size (61,936 bp) than the TAC APC10 (67,963 bp) and the amount of occurring deletions is highly correlated with the size of the TAC clone (Zhou *et al.*, 2011*b*), screening more T1s might have provided a good single-locus line for downstream analysis.

As sequencing of the TAC APC10 NGS^{yellow} revealed no mutations, the lack of complementation of the mutant phenotype is rather unexpected. Previous research showed that yeast *apc10* mutants could be complemented using the endogenous *APC10* sequence (Eloy *et al.*, 2011). Moreover, N-terminally TAP-tagged APC10 expressed in plant cell suspension cultures through the 35S promoter successfully delivered the full APC/C complex upon TAP purification (Van Leene *et al.*, 2007), and own data), showing the usability of the N-terminally tagged protein. It could be that the introduction of more genes than only the gene of interest within the TAC interferes with the expression of the tagged APC10 (Zhou *et al.*, 2011*a*), or that some *cis*-regulatory elements were still missing (Zhou *et al.*, 2011*a*). A back-up construct was made for APC10, combining both the endogenous promoter and gene and the NGS^{yellow} TAP tag through Gateway, but this construct also failed to complement the mutant phenotype of *apc10-2*. This might indicate that the tag interferes with the functionality of the protein, although again, we cannot exclude that the lack of necessary *cis*-regulatory elements for expression in the ovules might be the underlying reason.

USING TAC RECOMBINEERING FOR FUNCTIONAL PROTEOMICS IN PLANTS

Although recombineering of BACs for functional analysis and mutagenesis is a widely adapted and powerful method in many model organisms, such as mouse, zebrafish, *C. elegans* and *Drosophila* (Lee *et al.*, 2001; Swing and Sharan, 2004; Park *et al.*, 2004; Warming, 2005; Sarov *et al.*, 2006; Poser *et al.*, 2008; Venken *et al.*, 2008; Lee *et al.*, 2009), transfer of the technique to the plant field seems to lag behind. Nonetheless, several TAC and BIBAC (binary Bacterial Artificial Chromosomes, another type of plant transformation competent vectors) libraries for species, such as rice (Redei, 2004), (Belzile *et al.*, 2007), (Q *et al.*, 2002)), tobacco (Hamilton *et al.*, 1996), tomato (Hamilton *et al.*, 1999) and soybean (Wu *et al.*, 2009) have been constructed, but their usage seems to be rather limited.

In *Arabidopsis,* the transformation of these huge molecules through *Agrobacterium* is rather disadvantageous. Despite their ability to complement mutant phenotypes, their transfer often results in major deletions (Liu *et al.*, 1999; Pietra *et al.*, 2013). As a consequence, most groups that use the recombineering technique in plants subclone only a small portion of the BAC sequence (Bitrián *et al.*, 2011; (Rips *et al.*, 2014); (Han *et al.*, 2015); (Peret *et al.*, 2012). They focus mostly on the area around the GOI, and they subclone this part via, for example, gap repair into a vector optimised for *Agrobacterium* transfer. The resulting reduced size of the genomic region (5-30 kb) is easier to manipulate. Additionally, reducing the size eliminates the formation of aberrant phenotypes that are often associated with the insertion of all the other genes present on the vector (Zhou *et al.*, 2011*a*).

MATERIALS AND METHODS

Cloning of the GS^{rhino}_AMP and GS^{yellow}_AMP TAP tag recombineering cassettes

The sequences for the C- and N-terminal GS^{rhino} and GS^{yellow} tags were amplified from pUC57-attR2-C-GS-RhP-TAP-attL3 or pUC57-N-GS-RhP-TAP and pENR2L3_CGS^{yellow} or pENL1L2_NGS^{yellow} respectively (See Chapter 3, Materials and Methods). Primers used here were the Fwd + Rev TAC_C-GS$^{rhino/yellow}$ or Fwd + Rev TAC_N-GS^{rhino} (see Table S1), which also contain the necessary att sites for Gateway Cloning. The FRT_*AmpR_FRT* sequence was picked up from pJMA690 (provided by Hélène Boisivon) using Fwd_TAC_C-GS^{rhino}_FRT; Fwd_TAC_N-GS^{rhino}_FRT or Fwd_TAC_C-GS^{yellow}_FRT in combination with Rev_TAC_AMP_FRT primers. To allow the use of the universal primer sequences (which provide a standard linker between the tag and the bait), the same linker sequences (Zhou *et al.*, 2011*a*) were integrated 5′ and 3′ of the cassettes. The TAP tag sequences were then combined with the AmpR sequence through extension overlap PCR (Atanassov *et al.*, 2009), and cloned into pDONR221 through a BP reaction.

JatY clones and Recombineering

The following JatY clones were selected through the www.arabidopsislocalizome.org website, and ordered from the John Innes Centre (Norwich, UK): TOR: JAtY77P22; AN3: JAtY76I24; APC10: JatY54H07; AUR1: JAtY80G11; and TPLATE: JAtY73B08. Next, the TAP tag_AMP cassettes were amplified through PCR with gene-specific REC Fwd+Rev primers (See Table S1), and purified over gel. For the construction of the 200-bp homologous arm-containing cassettes, the 50-bp homologous arm cassette of AUR1 and TPLATE were used as a template for pick-up PCR with primers containing 150 bp extra. Isolation of the JAtY clones, transfer to the SW105 strain and recombineering was carried out as previously described (Warming, 2005; Zhou *et al.*, 2011*a*; Pietra *et al.*, 2013). Selection of transformed *E. coli* colonies was carried out on LB medium supplemented with 100 µg/mL carbenicillin and resulted in respectively ten, ten, four and four colonies harbouring the TAC of AN3, APC10, AUR1 or TPLATE. Induction of the *flippase* gene through treatment with L-arabinose was carried out as described (Lee *et al.*, 2001). All intermediate steps were verified through Colony PCR and sequencing with *AmpR* and gene-specific primers (see Table S1).

Transfer to Agrobacterium and transformation of *Arabidopsis*

Transfer of the TAP-tagged TAC DNA and the transformation of *Arabidopsis* via floral dip with Transfer of the TAP-tagged TAC DNA and the transformation of *Arabidopsis* via floral dip with the *Agrobacterium* strain C58C1 pMp90 was carried out as follows. Per construct 5 µL of vector DNA was added to 50 µL electro-competent *Agrobacterium* cells. Cells were electroporated in cold 0.1-cm electroporation cuvettes with a Micropulse Electroporator (Bio-Rad) at 1.8kV. After electroporation, the bacteria were recovered in 1 mL LB medium or 1 h at 32°C with

continuous shaking. After incubation, 50 µL was plated out on LB plates supplemented with 25 µg/mL kanamycin and grown for 2-3 days at 28°C. For plant transformation, the resulting *Agrobacterium* clones were precultured by inoculating them in 1 mL LB$^-$ without antibiotics. After 8-9 h, an additional 10 mL of LB$^-$ medium without antibiotics was added, and the culture was grown on a shaker overnight at 28 °C. The next day, 40 mL of MQ containing 5% glucose and 0.05% Silwet-77 was added to the culture and used directly to transform the plants. Plants were dipped for 10 s, and covered afterwards with saran wrap for 24 h.

Plant materials, growth conditions and genotyping PCR

The *apc10-2* mutant line was obtained from Nubia Eloy, and hemizygous plants were identified through PCR as described (Eloy *et al.*, 2011). Hemizygous T1 plants that carried the TAC APC10 NGS^{yellow} construct were selected on plates containing 1/2MS medium, supplemented with 1% (w/v) sucrose, 50 µg/mL kanamycin and 20 µg/mL L-phosphinothricin, or directly in soil through spraying (Logemann *et al.*, 2006) with Basta (Bayer, 560 µL/L). For the selection on plate we used the protocol provided by Dr. Alonso's group. The protocol wws as follows: sterilised seeds were first vernalised for 2 days in water and plated out on selection medium by suspending them with a 10-mL pipette in melted, pre-cooled 0.7% top agarose (1-3 volumes per volume of seed suspension). About 8,000 seeds (~0.2 g dry weight) were resuspended in 5-15 mL top agarose per 150-mm Petri plate containing 50-70 mL selection medium. Plates were then incubated in light for 1-2 h at room temperature and transferred to darkness for 3 days at 22°C. Finally, plates were transferred to light for 2-5 days and green transformants were selected.

Expression Analysis and TAP

The protocols for protein purification and the expression analysis of TOR-GS^{rhino} in seedlings as well as the subsequent TAP purifications are described in Chapter 3. Removal of unspecific proteins identified by MS was established by using an updated version of the published list of contaminants and unspecific proteins that often co-purify in TAPs (Van Leene *et al.*, 2014), which is described in Chapter 3.

REFERENCES

Adam G, Mullen JA, Kindle KL. 1997. Retrofitting YACs for direct DNA transfer into plant cells. The Plant Journal **11**, 1349–1358.

Atanassov II, Atanassov II, Etchells J, Turner SR. 2009. A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods **5**, 14–11.

Beauchamp EM, Platanias LC. 2013. The evolution of the TOR pathway and its role in cancer. Oncogene **32**, 3923–3932.

Bitrián M, Roodbarkelari F, Horváth M, Koncz C. 2011. BAC-recombineering for studying plant gene regulation: developmental control and cellular localization of SnRK1 kinase subunits.The Plant Journal **65**, 829–842.

Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling MC, Bottomley J, Brown SD, Bürger A, Bult CJ, Bushell W, Collins FS, Desaintes C, Doe B, Economides A, Eppig JT, Finnell RH, Fletcher C, Fray M, Frendewey D, Friedel RH, Grosveld FG, Hansen J, Hérault Y, Hicks G, Hörlein A, Houghton R, Hrabé de Angelis M, Huylebroeck D, Iyer V, de Jong PJ, Kadin JA, Kaloff C, Kennedy K, Koutsourakis M, Lloyd KC, Marschall S, Mason J, McKerlie C, McLeod MP, von Melchner H, Moore M, Mujica AO, Nagy A, Nefedov M, Nutter LM,

Pavlovic G, Peterson JL, Pollock J, Ramirez-Solis R, Rancourt DE, Raspa M, Remacle JE, Ringwald M, Rosen B, Rosenthal N, Rossant J, Ruiz Noppinger P, Ryder E, Schick JZ, Schnütgen F, Schofield P, Seisenberger C, Selloum M, Simpson EM, Skarnes WC, Smedley D, Stanford WL, Stewart AF, Stone K, Swan K, Tadepally H, Teboul L, Tocchini-Valentini GP, Valenzuela D, West AP, Yamamura K, Yoshinaga Y, Wurst W. 2012. The mammalian gene function resource: the International Knockout Mouse Consortium. Mammalian Genome **23**, 580-586

Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, Ahn J, Dewar-Darch D, Reguly T, Tang X, Almeida R, Qin ZS, Pawson T, Gingras AC, Nesvizhskii AI, Tyers M. 2010. A Global Protein Kinase and Phosphatase Interaction Network in Yeast. Science **328**, 1043–1046.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method forAgrobacterium-mediated transformation of*Arabidopsis* thaliana. The Plant Journal **16**, 735–743.

Collings DA. 2013. Subcellular localization of transiently expressed fluorescent fusion proteins. Methods Molecular Biology **1069**, 227–258.

Copeland NG, Jenkins NA, Court DL. 2001. Recombineering: a powerful new tool for mouse functional genomics. Nature Reviews Genetics **2**, 769–779.

Dedon PC, Soults JA, David Allis C, Gorovsky MA. 1991. A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. Analytical Biochemistry **197**, 83–90.

Deprost D, Yao L, Sormani R, Moreau M, Leterreux G, Nicolaï M, Bedu M, Robaglia C, Meyer C. 2007. The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. EMBO reports **8**, 864–870.

Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GTS, Inzé D, Ferreira PCG. 2011. The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. The Plant Journal **68**, 351–363.

Garcia N, Li Y, Dooner HK, Messing J. 2017. Maize defective kernelmutant generated by insertion of a Dselement in a gene encoding a highly conserved TTI2 cochaperone. Proceedings of the National Academy of Sciences **114**, 5165–5170.

Hamilton CM, Frary A, Lewis C, Tanksley SD. 1996. Stable transfer of intact high molecular weight DNA into plant chromosomes. Proceedings of the National Academy of Sciences **93**, 9975–9979.

Hamilton CM, Frary A, Xu Y, Tanksley SD, Zhang H-B. 1999. Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector. The Plant Journal **18**, 223–229.

Han SW, Alonso JM, Rojas-Pierce M. 2015. REGULATOR OF BULB BIOGENESIS1 (RBB1) Is Involved in Vacuole Bulb Formation in *Arabidopsis*. PL0S ONE **10**, e0125621.

Henriques R, Magyar Z, Bögre L. 2014. S6K1 and E2FB are in mutually antagonistic regulatory links controlling cell growth and proliferation in Arabidopsis. Plant Signaling & Behavior **8**, e24367.

Hirose Y, Suda K, , Liu YG, Sato S, Nakamura Y, Yokoyama K, Yamamoto N, Hanano S, Takita E, Sakurai N, Suzuki H, Nakamura Y, Kaneko T, Yano K, Tabata S, Shibata D. 2015. The *Arabidopsis* TAC Position Viewer: a high-resolution map of transformation-competent artificial chromosome (TAC) clones aligned with the *Arabidopsis thaliana* Columbia-0 genome. The Plant Journal **83**, 1114–1122.

John F, Roffler S, Wicker T, Ringli C. 2011. Plant TOR signaling components. Plant Signaling & Behavior **6**, 1700–1705.

Kaizuka T, Hara T, Oshiro N, Kikkawa U, Yonezawa K, Takehana K, Iemura S, Natsume T, Mizushima N. 2010. Tti1 and Tel2 are critical factors in mammalian target of rapamycin complex assembly. **285**, 20109–20116.

Kim SG, Hoffman GR, Poulogiannis G, Buel GR, Jang YJ, Lee KW, Kim BY, Erikson RL, Cantley LC, Choo AY, Blenis J. 2013. Metabolic stress controls mTORC1 lysosomal localization and dimerization by regulating the TTT-RUVBL1/2 complex. Mol Cell **49**, 172–185.

Lee DW, Lee S, Oh YJ, Hwang I. 2009. Multiple sequence motifs in the rubisco small subunit transit peptide independently contribute to Toc159-dependent import of proteins into chloroplasts. Plant Physiology **151**, 129–141.

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG. 2001. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics **73**, 56–65.

Li Y, Uhm T, Ren C, Wu C, Santos TS, Lee MK, Yan B, Santos F, Zhang A, Scheuring C, Sanchez A, Millena AC, Nguyen HT, Kou H, Liu D, Zhang HB. 2007. A plant-transformation-competent BIBAC/BAC-based map of rice for functional analysis and genetic engineering of its genomic sequence. Genome **50**, 278–288.

Liu P, Jenkins NA, Copeland NG. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Research **13**, 476–484.

Liu YG, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D. 1999. Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proceedings of the National Academy of Sciences **96**, 6535–6540.

Loewith R, Jacinto E, Wullschleger S, Lorberg A, Crespo JL, Bonenfant D, Oppliger W, Jenoe P, Hall MN. 2002. Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control. Mol Cell **10**, 457–468.

Logemann E, Birkenbihl RP, Ülker B, Somssich IE. 2006. An improved method for preparing Agrobacterium cells that simplifies the *Arabidopsis* transformation protocol. Plant Methods **2**, 16.

Mahfouz MM, Kim S, Delauney AJ, Verma DPS. 2006. *Arabidopsis* TARGET OF RAPAMYCIN interacts with RAPTOR, which regulates the activity of S6 kinase in response to osmotic stress signals. The Plant Cell **18**, 477–490.

Moreau M, Azzopardi M, Clément G, Dobrenel T, Marchive C, Renne C, Martin-Magniette ML, Taconnat L, Renou JP, Robaglia C, Meyer C. 2012. Mutations in the *Arabidopsis* Homolog of LST8/G L, a Partner of the Target of Rapamycin Kinase, Impair Plant Growth, Flowering, and Metabolic Adaptation to Long Days. Plant Cell **24**, 463–481.

Mosberg JA, Lajoie MJ, Church GM. 2010. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics **186**, 791–799.

Murphy KC. 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. Journal of Bacteriology **180**, 2063–2071.

Muyrers JP, Zhang Y, Testa G, Stewart AF. 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Research **27**, 1555–1557.

Nietzsche M, Landgraf R, Tohge T, Börnke F. 2016. A protein–protein interaction network linking the energy-sensor kinase SnRK1 to multiple signaling pathways in Arabidopsis thaliana. Current Plant Biology **5**, 36–44.

Park SW, Lin J, Ghosh B, Leach SD. 2004. Lambda Red-Mediated Bac Recombineering Allows Real-Time Visualisation Of Pancreas Development In Living Zebrafish Embryos. Pancreas **29**, 341.

Peret B, Swarup K, Ferguson A, Seth M, Yang Y, Dhondt S, James N, Casimiro I, Perry P, Syed A, Yang H, Reemmer J, Venison E, Howells C, Perez-Amador MA, Yun J, Alonso J, Beemster GT, Laplaze L, Murphy A, Bennett MJ, Nielsen E, Swarup R. 2012. AUX/LAX Genes Encode a Family of Auxin Influx Transporters That Perform Distinct Functions during *Arabidopsis* Development. Plant Cell **24**, 2874–2885.

P**eterson KR, Navas PA, Stamatoyannopoulos G**. 1998. beta-YAC transgenic mice for studying LCR function. Ann N Y Acad Sci **850**, 28–37

Pietra S, Gustavsson A, Kiefer C, Kalmbach L, Hörstedt P, Ikeda Y, Stepanova AN, Alonso JM. 2013. *Arabidopsis* SABRE and CLASP interact to stabilize cell division plane orientation and planar polarity. Nature Communications **4**, 2779.

Poser I, Sarov M, Hutchins JRA, Hériché JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA. 2008. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nature Methods **5**, 409–415.

Poteete AR, Volkert MR. 1988. Activation of recF-dependent recombination in Escherichia coli by bacteriophage lambda- and P22-encoded functions. Journal of Bacteriology **170**, 4379–4381.

Redei GP (Ed.). 2004. *Transformation-Competent Artificial Chromosome Vector (TAC)*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Rexin D, Meyer C, Robaglia C, Veit B. 2015. TOR signalling in plants. Biochem. J. **470** 1–14.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology **17**, 1030–1032.

Rips S, Bentley N, Jeong IS, Welch JL, Schaewen von A, Koiwa H. 2014. Multiple N-glycans cooperate in the subcellular targeting and functioning of *Arabidopsis* KORRIGAN1. **26**,The Plant Cell 3792–3808.

Sarbassov DD, Bulgakova O, Bersimbaev RI, Shaiken T. 2012. Isolation of the mTOR complexes by affinity purification. Methods Molecular Biology **821**, 59–74.

Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF. 2006. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nature Methods **3**, 839–844.

Sawitzke JA, Thomason LC, Costantino N, Bubunenko M, Datta S, Court DL. 2007.

Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. Methods Enzymology **421**, 171–199.

Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. Nature Protocols **4**, 206–223.

Swing DA, Sharan SK. 2004. BAC rescue: a tool for functional analysis of the mouse genome. Methods Mol Biol **256**, 183–198.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. Cell **33**, 25–35.

Takai H, Xie Y, de Lange T, Pavletich NP. 2010. Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes. Genes Dev. 2010 Sep **15**;24(18):2019-30.

Tao Q, Wang A, Zhang HB. 2002. One large-insert plant-transformation-competent BIBAC library and three BAC libraries of Japonica rice for genome research in rice and other grasses. Theoretical and Applied Genetics **105**, 1058–1066.

The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant*Arabidopsis* thaliana. Nature **408**, 796–815.

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. Nature Protocols **10**, 169–187.

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, Inzé D, Witters E, De Jaeger G. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis* thaliana. Molecular & Cellular Proteomics **6**, 1226–1238.

Venken KJT, Kasprowicz J, Kuenen S, Yan J, Hassan BA, Verstreken P. 2008. Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. Nucleic Acids Res **36**, e114–e114.

Wang J, Sarov M, Rientjes J, Fu J, Hollak H, Kranz H, Xie W, Stewart AF, Zhang Y. 2006. An improved recombineering approach by adding RecA to lambda Red recombination. Mol Biotechnol **32**, 43–53.

Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. 2005. Simple and highly efficient BAC recombineering using galK selection. Nucleic Acid Res **33**, e36–e36.

Wilson TE, Fahrner TJ, Johnston M, Milbrandt J. 1991. Identification of the DNA-binding site for NGFI-B by genetic selection in yeast. Science **252**, 1296–1300.

Wu C, Proestou D, Carter D, Nicholson E, Santos F, Zhao S, Zhang H-B, Goldsmith MR. 2009. Construction and sequence sampling of deep-coverage, large-insert BAC libraries for three model lepidopteran species. BMC Genomics **10**, 283.

Wu Q, Luo A, Zadrozny T, Sylvester A, Jackson D. 2013. Fluorescent protein marker lines in maize: generation and applications. The International journal of developmental biology **57**, 535–543.

Wullschleger S, Loewith R, Hall MN. 2006. TOR Signaling in Growth and Metabolism. Cell **124**, 471–484.

Wullschleger S, Loewith R, Oppliger W, Hall MN. 2005. Molecular Organization of Target of Rapamycin Complex 2. Journal of Biological Chemistry **280**, 30697–30704.

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

Young KH. 1998. Yeast two-hybrid: so many interactions,(in) so little time... Biology of reproduction **58** 302-311

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. Proceedings of the National Academy of Sciences **97**, 5978–5983.

Zhang Y, Buchholz F, Muyrers JP, Stewart AF. 1998. A new logic for DNA engineering using recombination in Escherichia coli. Nature Genetics **20**, 123–128.

Zhou L-Y, Jiang D-G, Wu H, Zhuang C-X, Liu Y-G, Mei M-T. 2005. Development of transformation system of rice based on transformation-competent artificial chromosome (TAC) vector. Acta genetica Sinica **32**, 514–518.

Zhou R. 2010. Recombineering-based Gene Tagging in *Arabidopsis*.

Zhou R, Benavente LM, Stepanova AN, Alonso JM. 2011. A recombineering-based gene tagging system for *Arabidopsis*. The Plant Journal **66**, 712–723.

SUPPLEMENTAL FIGURES

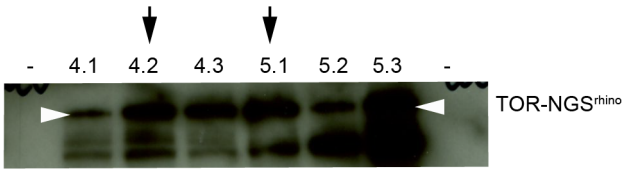


Figure S1. Western blot analysis of TAC TOR-GS^{rhino}-expressing seedlings. White arrowheads indicate TOR-GS^{rhino}; black arrows show selected lines for TAP; - : Mock (WT) control

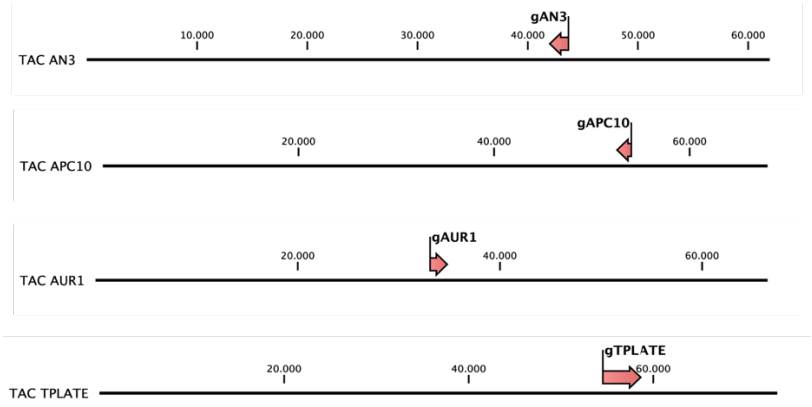


Figure S2. Overview of the TACs used for recombineering with the GS^{yellow} tag. Position of the gene of interest is indicated.

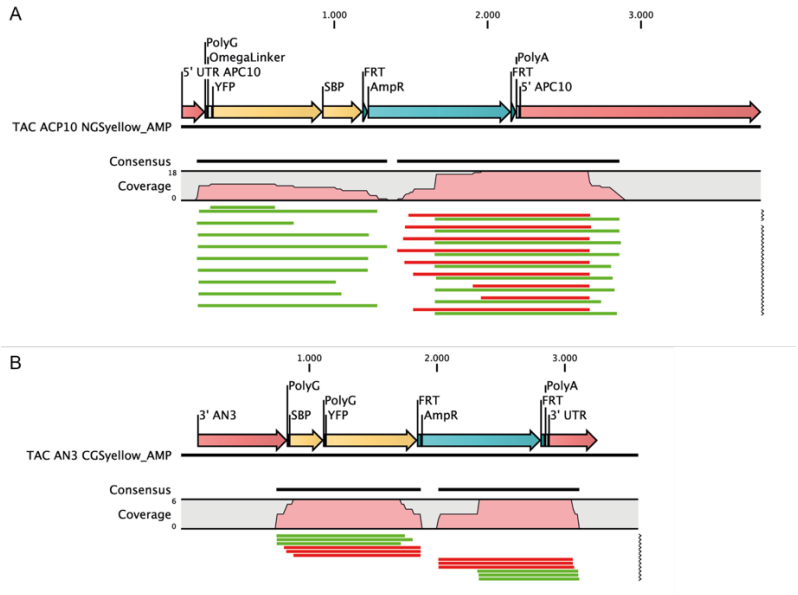


Figure S3. Allignement of sequencing results before flippase induction. A) TAC APC10 NGS^{yellow}_AMP (9 clones). **B)** TAC AN3 GS^{yellow}_AMP (3 clones). AmpR: Ampicilin Resistance Marker; SBP: Streptavidin Binding Peptide; YFP: Yellow Fluorescent Protein; FRT: Flippase Recognition Target; Sequencing results from forward primers in green, from reverse primers in red.

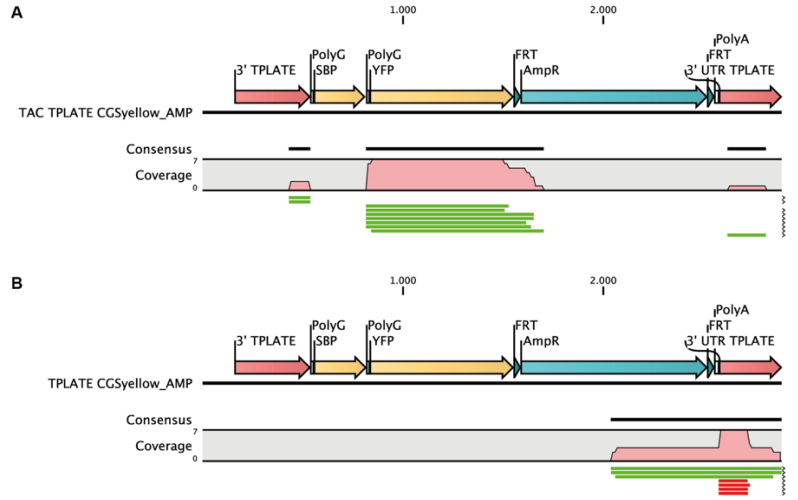


Figure S4. Alignement of sequencing results to TAC TPLATE GS^{yellow}_AMP. (A) The streptavidin-binding peptide (SBP) part of the GS^{yellow} tag is missing for the initial seven clones. **(B)** Correct sequence according to the forward primers (green), but the reverse primers (red) failed to amplify the AmpR (Ampicillin Resistance Marker) cassette (4 clones). YFP: Yellow Fluorescent Protein; FRT: Flippase Recognition Target.

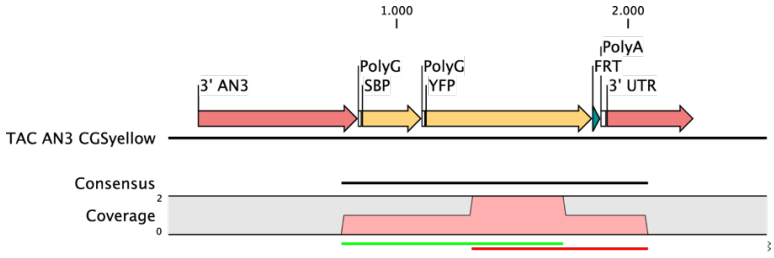
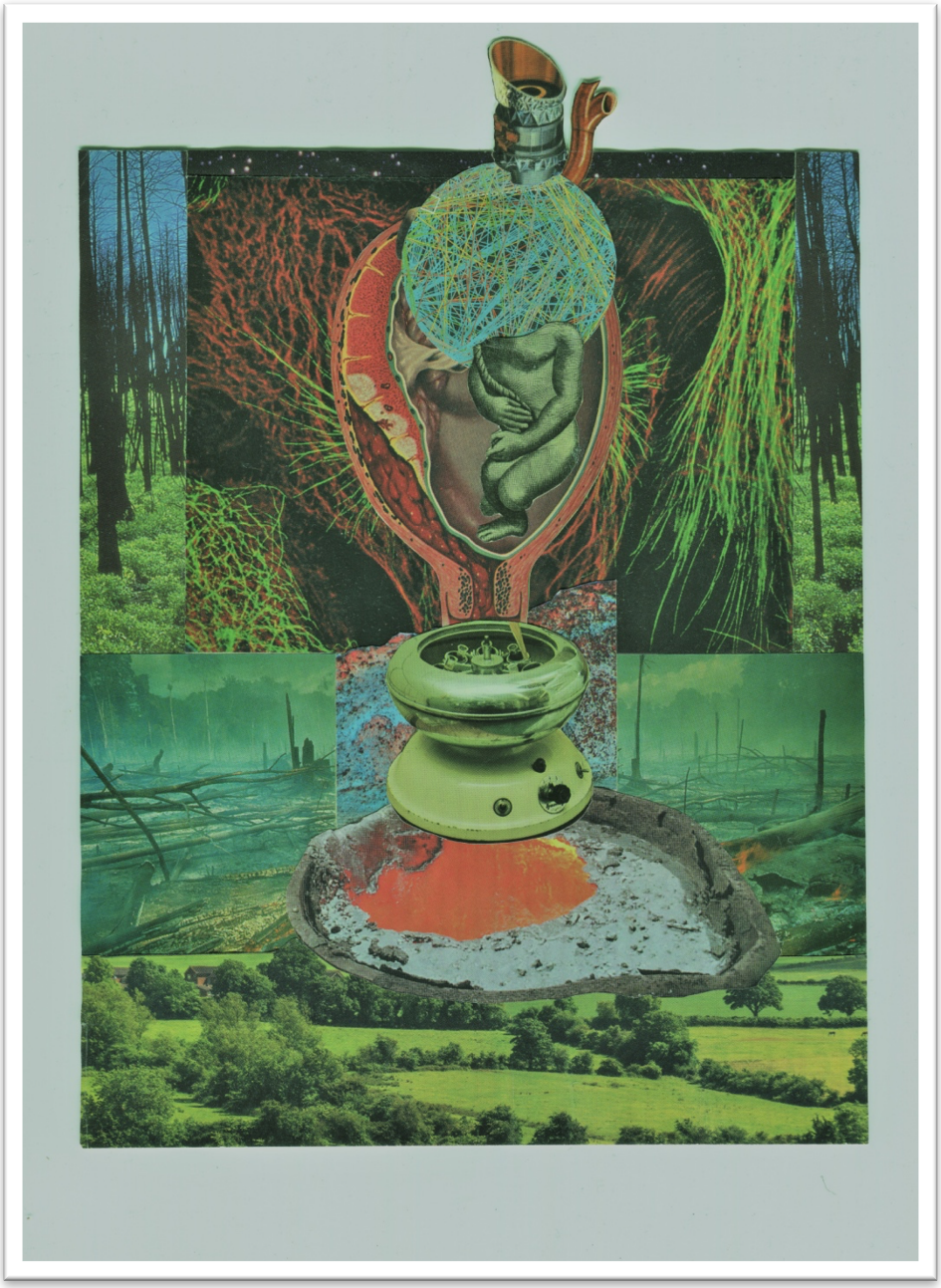


Figure S5. Allignement of sequencing results for TAC AN3 GS^{yellow} after flippase induction. Both the forward (green) as reverse (red) primer provided the correct sequence. AmpR: Ampicilin Resistance Marker; SBP: Streptavidin Binding Peptide; YFP: Yellow Fluorescent Protein; FRT: Flippase Recognition Target

Supplementary Table S1. Primers used for TAC recombineering

Function	Name	Sequence
Pick up Tag Cassette	**AN3_TAC_REC_Fwd**	CAGGGGATGGTGGAGAAACCCTTTACTTGAAATCATCAGATGA TGGGAATGGAGGTGGAGGTGGAGCT
	AN3_TAC_REC_Rev	TGCTGCTGCTGCTGCTTCATCATTAAGATTCCTTGTGCCACACTC TTTCAGGCCCCAGCGGCCGCAGCAGCACC
	APC10_N_REC_Fwd	ACAAATTGTTGGAATCTCAATTCTCGAATCTGACATCTCGGAGA AGAGATGgaggtggaggtggagct
	APC10_N_REC_Rev	TTTCCTCCACTAATCTTTCCTTCTTCTTCCGATTCCGATGACTCTG TCGCGGCCCCAGCGGCCGCAGCAGCACC
	AUR1_200_REC_Fwd	TGCTAGTTATTGAATCTGAAGTCTCACATATTTTCTCAATTGTAT GTATAGATGCTTGTCAAGGAGTCTTCGCAACGTCTGCCATTGCA CAAGCTTCTGG**AGCATCCGTGGATCGTGC**AAAACGCTGATCCTT CTGGAATCTACAGAGTT*ggtggcggtggaggctcaATGGACGAGAAGA CCACCGGCTGGAGAGGCGG*
	AUR1_REC_Fwd	AGCATCCGTGGATCGTGCAAAACGCTGATCCTTCTGGAATCTAC AGAGTTGGAGGTGGAGGTGGAGCT
	AUR1_REC_Rev	GATAACAGTTTTGGTCGAGATTGGCGAGAACAGTAAGCGTTTT GTTTTTAGGCCCCAGCGGCCGCAGCAGCACC
	TOR_TAC_REC_Fwd	CAGATTTAGGGTTTTTGGAAGATTTTTGTGGCTGGACTGGCCCT GCAGCGGGAGGTGGAGGTGGAGCT
	TOR_TAC_REC_Rev	GAAGCCATGGATGCAGGCCGTCCAGCCACAAAAGATTGCGACG AGGTAGAGGCCCCAGCGGCCGCAGCAGCACC
	TPLATE_200_REC_Fwd	AAGGACAAGGAAGAGAAGAAGAAAAAAGAGAAGGAAAAAGGA ACATTTTCTAAGCTGACAGCAGAAGAAACAGAGCACATGGCGC TTCAGGCAGCAGTGC**TCCAAGAATGGCACATCTTG**TGCAAAGA TAGAAAATATACCAAAGTTAAC*ggtggcggtggaggctcaATGGACGA GAAGACCACCGGCTGGAGAGGCGG*
	TPLATE_REC_Fwd	TCCAAGAATGGCACATCTTGTGCAAAGATAGAAAATATACCAAA GTTAACGGAGGTGGAGGTGGAGCT
	TPLATE_REC_Rev	AAACAAAGATTGTTGTAGCTTCATCTGAGAAAAAAGTACTCTCG CTCTTAGGCCCCAGCGGCCGCAGCAGCACC
Pick up AMP	**Fwd_TAC_C-GSrhino_FRT**	ACTGTCACTGAATAGGGGAAGTTCCTATACTTTCTAGAGAATAG GAACTTCTTCAAATATG
	Fwd_TAC_C-GSyellow_FRT	GAGCTGTACAAGtgaGGGAAGTTCCTATACTTTCTAGAGAATAG GAACTTCTTCAAATATG
	Fwd_TAC_N-GSrhino_FRT	GGGCAGCGCGAGCCGGGGAAGTTCCTATACTTTCTAGAGAATA GGAACTTCTTCAAATATG
	Rev_TAC_AMP_FRT	GGGGACCACTTTGTACAAGAAAGCTGGGTCGGCCCCAGCGGCC GCAG
Pick up tag	**Fwd_TAC_C-GSrhino**	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGGAGGTGGAGGT GGAGCTGAGCAGAAGCTTATCTCCGAGG

	Fwd_TAC_C-Gsyellow	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGAGGTGGAGGTG GAGCTATGGACGAGAAGACCACCGG
	Fwd_TAC_N-GSrhino	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGGAGGTGGAGGT GGAGCTaacaacaaacaacaaacaacattacaattacaatgg
	Rev_TAC_C-GSrhino	GTATAGGAACTTCCCCTATTCAGTGACAGTGAAAGTCTTTGTC
	Rev_TAC_C-GSyellow	GTATAGGAACTTCCCTCACTTGTACAGCTCGTCCATGC
	Rev_TAC_N-GSrhino	GTATAGGAACTTCCCCGGCTCGCGCTGCCCCTG
Test Primers	**AN3_TAC_TEST_Fwd**	CTGGTGGGTTTCATGATTTTG
	AN3_TAC_TEST_Rev	ATTACGACACAACTTGGAGCC
	APC10_TEST_Fwd	TTTCAAAACCTCGTCGGAGA
	APC10_TEST_Rev	TGGAGGTCTTGGTTGATAAGC
	AUR1_TEST_Fwd	TTCCCTCCCAAACCAATAATC
	AUR1_TEST_Rev	TTTCCTTCTTCCGTTAATGCC
	GSyellow_TAC_intern_Rev	CCTCCTTGAAGTCGATGCCCTT
	TAC_Amp_Fwd	TTACGGATGGCATGACAGTAAG
	TAC_Amp_Rev	TTATTGTCTCATGAGCGGATACA
	TAC_NGSrhino_intern_Fwd	GGAGTCTGGACCTACGATGA
	TAC_NGSrhino_intern_Rev	TCACAGTGAATGTCTTCGTG
	TOR_TAC_TEST_Fwd	TGATTTCTTCTGCGATCTTCG
	TOR_TAC_TEST_Rev	ACCTTAGGATTACCACGGCTG
	TPLATE_TEST_Fwd	GAGGGAAAGGACAAGGAAGAG
	TPLATE_TEST_Rev	TGCATCCTTGGTTCCATAAAC
Test TAC *Arabidopsis*	**AN3_TAC_LB_Rev**	ACCGAAAAAGATAACTCTGGCGTAGAGAG
	AN3_TAC_RB_Fwd	ACTATCCGCTAGAGAGAGCCA
	APC10_TAC_LB_Rev	CACCATGCCGTGGATAGTCCAT
	APC10_TAC_RB_Fwd	CCGGTGATTAAAGACTTCAACAGAGC
	pYLTAC7_LB_Fwd	CTAGGGATAACAGGGTAATGCGGC
	pYLTAC7_RB_Fwd	CGGTACCTTTGACATTCTAGGACTATATTG
	TOR_TAC_LB_Rev	AATGAAGCCAGGCATTGCAGC
	TOR_TAC_RB_Rev	CTCCTCCTACATCAATATTTTACAGGAAGG



André

PULL-DOWN VERSUS TAP-MS: DO YOU NEED TWO TO TANGO?

Nienke Besbrugge[1,2], Dominique Eeckhout[1,2], Nancy De Winne[1,2], Geert Persiau[1,2], Eveline Van De Slijke[1,2], Daniel Van Damme[1,2], Mieke Van Lijsebettens[1,2], Hilde Nelissen[1,2], Jelle Van Leene[1,2] and Geert De Jaeger[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2]Centre for Plant Systems Biology, VIB, 9052 Ghent, Belgium

Author contributions

G.D.J., H.N. and J.V.L. designed the research; N.B. performed most of the experiments and wrote the chapter; D.E., N.D.W., E.V.D.S. and G.P. carried out experiments and provided technical assistance; D.V.D. and M.V.L. provided plant material; J.V.L, H.N and G.D.J. supervised and complemented the writing

ABSTRACT

Biological processes are orchestrated by a myriad of proteins that interact with each other and with other macromolecules. Hence, identifying a protein's interaction partners in a proteome-wide manner through affinity purification coupled to mass spectrometry (AP-MS) is an important approach to unravel a protein's cellular functions. To help the identification of *bona fide* interactors in AP-MS, two different strategies can be followed: combining pull-down with label-free quantification of MS data to screen for enriched proteins, or addition of a second purification round together with the implementation of a tandem affinity purification (TAP) tag. Here, pull-downs and TAP experiments were performed on the same biological material in *Arabidopsis thaliana* cell cultures and seedlings, as well as in different developmental tissues of *Zea mays* leaves to compare both methods. For most baits, TAP perfomed better than GFP pull-down. However, GFP pull-downs are more advisable when biological input material is limited. The results thus reinforce the advantage of implementing a versatile tag, such as the GS^{yellow} TAP tag that allows the choice between both AP-MS approaches in light of the anticipated results.

INTRODUCTION

Proteins are one of the main cellular molecules that orchestrate a variety of complex biological processes. Although some proteins function in an independent manner, most of them require the interaction of other proteins for their activity. Given their time and space specific expression profiles and the differences in abundance and affinity of proteins, such interactions are very dynamic in nature and offer an enormous versatility and adaptability to each cell. Studying the complexity of PPIs and the underlying patterns has therefore become of major importance. On the one hand, for unknown proteins, identification of interaction partners for which the function is well known allows the extrapolation of the function to the protein under study, a principle better known as guilt-by-association. On the other hand, PPI analysis can also unravel new unique functional roles of well-studied proteins. Over the years, a variety of PPI discovering techniques have been developed, of which affinity purification coupled to mass spectrometry (AP-MS) is the most widely adopted to discover protein complexes *in situ* (Dedecker *et al.*, 2015). In general, AP uses a solid support, such as agarose or magnetic beads to capture the bait protein and its associated partners, providing a purified sample that can be analysed directly by MS. To circumvent the need for bait specific antibodies for the purification, genetic tagging of the protein of interest is commonly applied. Because of the nature of the AP procedure, a number of false positive or unspecific interactors, which will bind to the affinity matrix or the bait in an unspecific way, are always co-purified in the sample. Identifying these mostly very abundant and sticky proteins, and removing them from the list of co-purified proteins is very crucial, but difficult. In an attempt to help on this matter, a contaminant repository for affinity purification (CRAPome) has been constructed that contains a list of all proteins that were identified in control and mock experiments (Mellacheruvu *et al.*, 2013). The

idea behind it is that non-specific interactors that bind to the solid support or the tag will be consistent across purifications, and that using such a subtraction list can aid to distinguish the true interactors from background noise. The database, however, consists mostly of yeast and human experiments, making it less usable for AP-MS analysis in plants. Moreover, the nature of both the captured interactions as well as the contaminants is influenced by the choice of tag and tag position, the used conditions for protein extraction and the purification strategy (Keilhauer *et al.*, 2014).

Bona fide interactors can also be identified by using the quantitative data obtained through MS, a technique that is driven by the continuous increase in MS detector sensitivity (Meyer and Selbach, 2015). Based on the idea that these interactors will be more enriched in the sample when compared to control extracts, the change in their abundance ratio can be used as a measure for the interaction. Due to the low cost and the availability of well performing algorithms, Label Free Quantification (LFQ) methods are now becoming more and more popular, although stable isotope-based methods are more sensitive to changes in protein abundance (Mittler *et al.*, 2008). In plants, however, incorporation of these isotopes is less evident due to their autotrophic nature (Dedecker *et al.*, 2015). LFQ methods, such as the peptide ion counting MaxQuant algorithm, require a set of repeat experiments as well as statistical data analysis. In MaxQuant relative protein concentrations are determined by comparing the peak areas (calculated as the sum of ion counts over peptide elution time) of the protein peptides across the analysed eluate. Because the concentration is calculated in relation to the background proteins, the method requires 'dirty' control samples and is therefore a powerful tool to combine with single-step AP-MS (hereafter named 'pull-down') (Pardo and Choudhary, 2012).

Another strategy is to reduce the number of contaminants by using dual affinity tags that consist of two affinity modules separated by a restriction site, and allowing tandem affinity purifications (TAP). Addition of this second purification step causes loss of unspecific binding products that are linked to the first part of the tag or to the solid support that is used during the first round of purification. Indeed, TAP has been shown to deliver rather pure protein complexes that can be unravelled using MS, and it has been successfully applied in the analysis of the yeast protein interactome (Gavin *et al.*, 2002), as well as the cell cycle interactome in *Arabidopsis* cell cultures (Van Leene *et al.*, 2011). The protocol, originally established in yeast, has also been optimised for several plant species and tissues in both monocots and dicots (Rohila *et al.*, 2006; Van Leene *et al.*, 2014; Nelissen *et al.*, 2015; Dedecker *et al.*, 2016; Goossens *et al.*, 2016). Because most of the abundant background contaminants are absent in the eluate, the technique even allows the identification of substoichiometric interactions, as was shown by picking up the regulatory proteins of the APC/C complex through TAP on SAMBA (Eloy *et al.*, 2012). However, because of the stringent wash steps and the length of the purification protocol, transient and weak interactions are often lost during TAP. Combining TAP with LFQ methods is also a possibility, and recently published work

in *Zea mays* shows that it can capture dynamic changes in interaction partners by focussing on different tissues of a developing leaf (Nelissen *et al.*, 2015).

In plants, the availability of GFP reporter lines for visualisation and the development of excellent anti-GFP affinity resins (Rothbauer *et al.*, 2008) has caused an increase in pull-down experiments (Dedecker *et al.*, 2015), which is generally combined with LFQ using a WT or a mock GFP line as control. To provide a better insight into the differences gathered with pull-down versus TAP, we compared a standard plant pull-down protocol (Wendrich *et al.*, 2017) with our TAP protocol (Van Leene *et al.*, 2014) in both *Arabidopsis* and *Zea mays*. . To this end, a set of well-studied proteins (SAMBA, AUR1, AN3 and ZmAN3; see also Chapter 2) involved mainly in cell proliferation was selected.

SAMBA is a regulator of the plant-specific APC/C involved in the transition of metaphase-to-anaphase during cell cycle progression. By binding to A-type cyclins during early development of *Arabidopsis* seedlings, it helps to degrade the latter proteins. As such, cell cycle progression is stopped and plant growth is inhibited. Previous identification of interactors of SAMBA through TAP in *Arabidopsis* cell cultures revealed a total of 12 interacting proteins, most of them being subunits and regulators of the APC/C complex (Eloy *et al.*, 2012).

The second protein, AUR1, is a kinase with a major role in spindle pole formation. Together with its redundant homologue AUR2, it was shown to be expressed in dividing cells (Van Damme *et al.*, 2012), where they form a heterodimer. For its activity, AUR1 requires an interaction with the plant homologue of TPX2 (targeting protein for *Xenopus* kinesin-like protein 2), a microtubule nucleation factor. This association helps AUR1 to localise to the spindle microtubuli, and stimulates the auto-phosphorylation and activation of TPX2.

Finally, AN3 and its orthologue ZmAN3 are known transcriptional co-activators with a plant growth-regulatory function. Overexpression of AN3 in *Arabidopsis* was shown to extend cell proliferation before transition to cell expansion, which causes a significant increase in leaf growth (Vercruyssen *et al.*, 2014). Previous work in both *Arabidopsis thaliana* and *Zea mays* further described the association of (Zm)AN3 with subunits of the SWI/SNF chromatin-remodelling complex (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015), as well as interactions with transcription factors (TFs), such as the Growth Regulating Factors (GRFs) and WOX13 (Nelissen *et al.*, 2015). By zooming in on specific tissue in the leaf growth zone, TAP on ZmAN3 in combination with MaxLFQ revealed how these interactions with GRFs are dynamic along the growth zone (Nelissen *et al.*, 2015).

In Chapter 3, we demonstrated the use of the $GS^{green/yellow}$ TAP tag, which combines GFP/cYFP with a streptavidin binding peptide (SBP) for the purification of protein complexes, through TAP in both *Arabidopsis* cell cultures and seedlings, as well as in *Zea mays*. Primary binding with this tag occurs against the GFP/cYFP part, whereafter which the complexes can be eluted through enzymatic digest. The SBP part in the tag can then be used for a second round of purification, followed by final elution with biotin. To allow comparison of the pull-down and TAP approaches,

the selected baits were tagged with the GS$^{green/yellow}$TAP-tag. Both approaches can thus be carried out on the same transgenic line. Our results show that both AP techniques sometimes provide different sets of interactors for the same bait, and finally reveal the advantage of having both approaches at hand in light of the anticipated results.

RESULTS

PULL-DOWN VERSUS TAP-MS IN *ARABIDOPSIS* CELL CULTURES

To provide an insight into the differences gathered with a pull-down and a TAP, two protocols were selected that are GFP-based (Van Leene *et al.*, 2014; Wendrich *et al.*, 2017). Because of the in-house experience and the relative ease and speed of biomass production, a first comparison was carried out using *Arabidopsis* cell suspension cultures. Here, SAMBA was selected as bait protein and tagged with the GS^{green} TAP tag (See Chapter 3). Based on previous SAMBA studies, the tag was added at the C terminus (Eloy *et al.*, 2012). To enhance competition with the endogenous counterpart for incorporation into the complex, a strong constitutive promoter (35S) was selected (see Western blot analysis of expression in Figure S1). For each bait, duplicate TAP experiments were compared with triple pull-downs using an Orbitrap Velos MS (Van Leene *et al.*, 2014). Both protocols were performed on biological material coming from a single cell culture and analysed in a single run on the MS.

Table 1. Comparison of the interaction partners of SAMBA in *Arabidopsis* cell cultures as identified by pull-down or TAP-MS

SAMBA		Pull-down*		TAP*	Previously reported
ProteinID	Protein name	Difference[1]	-LogP[2]		
AT5G05560	APC1	8.027	3.841	2	x
AT2G04660	APC2			2	x
AT3G16320	APC3a/CDC27a	10.492	5.24	2	x
AT2G20000	APC3b/CDC27b/HBT (HOBBIT)	11.528	4.75	2	x
AT4G21530	APC4	7.325	2.80	2	x
AT1G06590	APC5	7.427	3.60	2	x
AT1G78770	APC6/CDC16	7.925	3.51	2	x
AT2G39090	APC7	11.737	6.40	2	x
AT3G48150	APC8/CDC23	10.002	4.15	2	x
AT2G18290	APC10	6.894	2.80	2	x
AT4G22910	CCS52A1			2	x
AT5G13840	CCS52B			2	x
AT4G33270	CDC20-1			2	x
AT1G35580	Cytosolic Invertase 1	8.114	2.23	2	
AT4G00755	F-box Family Protein			2	
AT1G32310	SAMBA (bait)	15.166	6.10	2	x
AT3G57860	UVI4-like	3.724	2.51	2	x

* Material originated from a single transgenic culture. Samples analysed in a single run

[1]Difference: $\log_2$ (LFQ ratio (SAMBA-GS^{green}/WT)) = $\log_2$ average LFQ intensity (group1, SAMBA-GS^{green}) - $\log_2$ average LFQ intensity (group2, WT); calculated from the Proteingroups file (Supplemental Data S1B) with the Perseus software (as described in the methods section; briefly: LFQ intensities were logged; missing values were imputed from a normal distribution around the detection limit of the mass spectrometer; average $\log_2$ LFQ intensity was calculated for each identification in both groups: SAMBA (green columns) and WT (PSB-L, yellow columns); for each identification, the difference in average $\log_2$ LFQ intensity between the two groups is calculated and shown here.

[2]-LogP: Test statistic: $-\log_{10}$ p-value, taking into account FDR=0.01 and S0=0.9

Removal of sticky and unspecific binders from the TAP datasets was carried out using an extended version of the previously described list of commonly co-purified proteins inherent to the purification protocol (Van Leene *et al.*, 2014) (see Chapter 3). Sticky and unspecific proteins were identified across 652 control experiments on extracts coming from WT (mock) and 127 different TAP-tagged bait cultures, including GUS-GS^{yellow}. For the pull-downs, significantly enriched interactors were identified from triplicate samples through standard settings based on MaxLFQ using a WT (mock) culture as control (Wendrich *et al.*, 2017). All MS data are presented in Supplemental Data S1. TAP-MS on SAMBA-GS^{green} provided interactions of SAMBA with 16 different proteins, of which ten (APC1, APC2, APC3a/CDC27a, APC3b/HBT/CDC27b, APC4, APC5, APC6, APC7, APC8 and APC10) belong to the APC/C core complex (Table 1). In addition, TAP also found the UVI4-like protein, a known interactor and negative regulator of the APC/C complex (Eloy *et al.*, 2012), as well as three activators: CCS52A1, CCS52B (homologues of Cdh1) and CDC20-1 (Eloy *et al.*, 2015). Two other proteins were also found, cytosolic invertase1 (AT1G35580) and an F-box family protein (AT4G00755). Pull-down on SAMBA, on the other hand, confirms the interactions with the APC/C core complex (except APC2), but fails to identify any of the APC-regulating proteins (see Table 1).

All interactors identified with pull-down for which no confirmation was found through TAP (66 in total) are presented in Table S1. Besides many known unspecifically associating interactors that are often found in TAPs (indicated in red in Supplemental Data S1B), there are still a lot of co-purifying proteins, such as, for example, the SKU5 similar protein 17 and 6 without reported interaction with SAMBA, which could interfere with the detection of low abundant regulators.

PULL-DOWN VERSUS TAP-MS IN *ARABIDOPSIS* SEEDLINGS

The results in cell cultures only show possible interactions in undifferentiated cells, but lack the developmental window or location of interaction. As such, the experimental value of both techniques should also be assessed in a more biologically relevant background, such as *Arabidopsis* seedlings or plant organs (see below). The increase in detection limit of MS for low abundant proteins, plant proteomics analysis is now switching to a more biological situation in which endogenous promoters and regulatory sequences are used for functional analysis. We chose to study AUR1 based on its specific expression profile during development. Given the relatively high number of dividing cells in the developing tissues of seedlings (Vercruyssen *et al.*, 2014), we anticipated the usability of AUR1 as a bait for functional analysis in such a set-up. Moreover, plants that have a mutation in both the AUR1 and AUR2 gene have a distinct phenotype compared to the wild-type. Previous data showed that this mutant phenotype can be complemented through the use of the AUR1-coding sequence coupled to the endogenous promoter (Van Damme *et al.*, 2012), and as such allows testing for interference of the added tag with the bait's function.

Because not much is known about the other interaction partners of AUR1 in *Arabidopsis*, a first affinity purification experiment was carried out in plant cell

cultures. Differences in interactors that originate from the choice of the TAP tag were neutralised by producing two different cultures, containing either AUR1-GS^{rhino} or AUR1-GS^{yellow}. Figure S1 shows a Western blot analysis of their respective accumulation levels in cell cultures. Based on previous tagged AUR1-GFP data (Van Damme *et al.*, 2012), both tags were fused at the C terminus of AUR1.

Table 2. Interaction partners of AUR1 in *Arabidopsis* as identified through TAP and pull-down in cell cultures and seedlings

AUR1		TAP		Pull-down		
				Difference[1]	-LogP[2]	
		Cell culture (35S)*		Seedlings (pAUR1)		
ProteinID	Protein name	GS^{rhino}	GS^{yellow}			
AT4G32830	AURORA1 (bait)	2	2	1	7.25	2.36
AT2G25880	AURORA2				5.49	3.32
AT1G03780	TPX2		1			
AT4G11990	TPXL2	2	2			
AT4G22860	TPXL3	2	2	1		
AT5G54740	SESA5 (seed storage albumin5)				4.17	3.97
AT1G79930	HSP91 (heat shock protein 91)				4.29	3.34

* Technical repeats

Used promoters are indicated between parentheses.

[1]Difference: $\log_2$ (LFQ ratio (AUR1-GS^{yellow}/WT)) = $\log_2$ average LFQ intensity (group1, AUR1-GS^{yellow}) - $\log_2$ average LFQ intensity (group2, WT); calculated from the Proteingroups file (Supplemental Data S2C) with the Perseus software (as described in the methods section; briefly: LFQ intensities were logged; missing values were imputated from a normal distribution around the detection limit of the mass spectrometer; average $\log_2$ LFQ intensity was calculated for each identification in both groups: AUR1 (green columns) and WT (Col-0, yellow columns); for each identification, the difference in average $\log_2$ LFQ intensity between the two groups is calculated and shown here.

[2]-LogP: Test statistic: $-\log_{10}$ p-value, taking into account FDR=0.05 and S0=0.1

Duplicate TAP experiments on the cultures rendered a list of two interactors that were found in each sample with both tags after filtering for false positives and unspecific interactors (see Table 2). These proteins, (AT4G11990 and AT4G22860, also known as TPXL2 and TPXL3 (TPX2-like 2 and 3)) have a specific TPX2-like signature and were previously predicted to interact with AUR1 based on the presence of the AUR1 interaction domain that is also present in TPX2 (Tomaštíková *et al.*, 2015). Moreover, interaction of TPXL2 and AUR1 was further confirmed through Y2H (see Figure 1A). TAP with the GS^{yellow} tag also confirmed the known interaction between AUR1 and TPX2, although only found once in two experiments.

To validate the interactions that were found in cell cultures in a more endogenous context, two independent single-locus transgenic *Arabidopsis* lines were made that express AUR1-GS^{yellow} from its endogenous promoter. To be able to test for interference of the tag with the function of the fusion protein and to eliminate competition of the bait with the endogenous counterpart for complex assembly, the *aur1-2aur2-2* double mutant (which originated from a combination of the SALK_031697 line with the Gabi-Kat line GK403B02) was used. The bushy, short internoded phenotype of this mutant was found to be converted back to wild-type in the transgenic lines. Also, the primary root length of *aur1-2aur2-2*, which is on average ~72% of the wild-type root length, was restored to WT levels in 12-day-old *aur1-2aur2-2* AUR1-GS^{yellow}-transformed plants (Figure 1B,C).

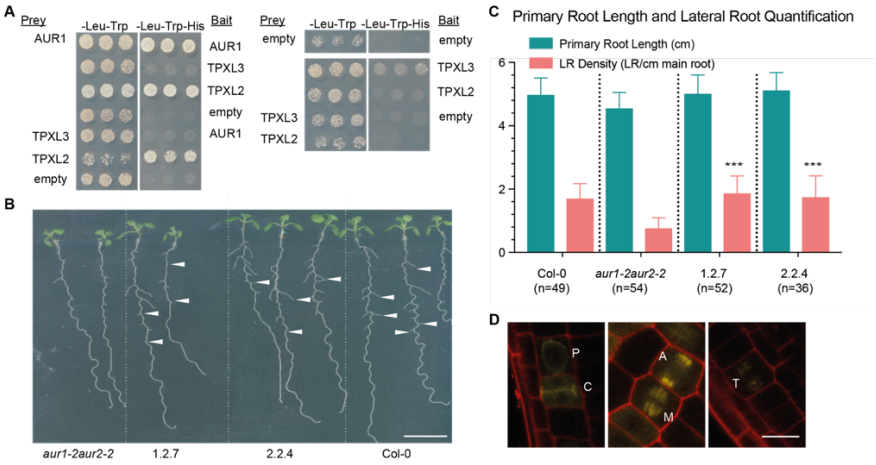


Figure 1. Functional analysis of AUR1. (A) Confirmation of binary interaction between AUR1 and TPXL2 and TPXL3 through Y2H. **(B)** Representative images of the seedlings used for quantification in **(C)** showing the rescue of the mutant phenotype. Arrowheads indicate lateral roots. Bar = 1 cm. **(C)** Quantification of lateral root density (11 DAS) comparing the *aur1-2aur2-2* double mutant (n = 54), two transgenic *aur1-2aur2-2* double mutant lines transformed with the open reading frame of AUR1 tagged with GS^{yellow} expressed from the *AUR1* promoter (n= 52 for line 1.2.7 and n=36 for line 2.2.4), and the wild-type (Col-0, n = 49). Lateral root densities are restored to wild-type levels, indicating full functionality of the tagged AUR1 (*t* test; triple asterisk: p<0.0001). Error bars represent SE (n=3). **(D)** Confocal images of hypocotyl cells (3 DAS) from the *aur1-2aur2-2* double mutant seedlings transformed with pAUR1:AUR1-GS^{yellow} used in **(C).** AUR1 localises to the mitotic spindle pole during the different stages of cell division. Cells were stained with propidium iodide. P: Prophase; C: Cytokinesis; A: Anaphase; M: Metaphase; T: Telophase. Bar = 10 µm.

Moreover, the decreased level of lateral root emergence in the double mutant, which is caused by a defect in division plane positioning during formative cell divisions, was also restored to WT levels (*t*-test, p<0.001) (Figure 1C). Next, we assessed whether the AUR1-GS^{yellow} fusion protein was able to migrate to its reported developmentally specific subcellular localisation, *i.e.*, the mitotic spindle pole during cell division (Van Damme *et al.*, 2012), by subjecting the roots of young seedlings to confocal microscopy (see Figure 1D). Movement of AUR1-GS^{yellow} during mitosis could be followed step by step: Localisation with the pre-prophase nucleus, association of AUR1 with the mitotic spindle during metaphase and anaphase, and final accumulation in the phragmoplast area and new cell plate during late telophase. Next, the transgenic line with the highest accumulation of the bait was determined through Western blot. Given that AUR1 is mostly present in dividing cells and TAP requires a lot of input material, it was crucial to also determine the timing when the ratio of dividing cells per weight was the highest. Hence, an immunoprecipitation of the bait was carried out on extracts obtained from seedlings from two different transgenic lines that were harvested at several time points (see Supplemental Figure S1B) The resulting selected line was then grown for 4 days after stratification (DAS) and subjected to duplicate TAP-MS/MS experiments using a Q-Exactive Mass Spectrometer for detection. Only two

proteins were found to be present in one of two technical repeats: AUR1 itself and TPXL3 (see Table 2 and Supplemental Data S2). Although this confirms the interaction of AUR1 and TPXL3 that was found in the cell culture experiment, the abundance of the AUR1-GS^{yellow} protein when using its endogenous promoter is clearly lower in seedlings. In this case, the low and time- and space-specific expression profiles make it more difficult to detect associating complexes through TAP in whole seedling extracts.

In an attempt to reduce the loss of transiently and weakly interacting proteins during TAP, we then performed a pull-down experiment (see Table 2 and Supplemental Data S2). As pull-down requires less biological input material than TAP, the ratio of dividing cells could be increased by performing these experiments on 3-day-old seedlings. Four proteins (AUR1, AUR2, HSP91 and SESA5) were found to be significantly enriched in the AUR1-GS^{yellow} pull-down sample (Figure S2).

One way to circumvent the problem of low native expression is by enriching cells that contain the bait's complexes by using, for instance, synchronised cell suspension cultures or specific tissues. Another option is to use a strong overexpression promoter, such as the 35S promoter that is commonly applied in plant proteomics studies. Such a strong promoter is often a necessity when performing AP-MS in a WT background to ensure proper competition with the endogenous non-tagged counterpart of the bait for complex formation. However, overexpression might also lead to false positive interactors due to ectopic expression of the bait.

Table 3. Interaction partners of AN3 as identified with pull-down in *Arabidopsis* seedlings

AN3		**Pull-down**		**Previously Reported**
ProteinID	**Protein name**	**Difference**	**-LogP**	
AT3G60830	ARP7	6.64	4.60	x
AT2G46020	BRM	10.68	3.82	x
AT4G18480	CHLI1	3.10	1.55	x
AT4G17330	G2482-1	6.25	4.02	x
AT1G21700	SWI3C	3.03	3.45	x
AT5G14170	SWP73B	5.29	4.23	X

[1]Difference: $\log_2$ (LFQ ratio (AN3-GS^{yellow}/WT)) = $\log_2$ average LFQ intensity (group1, AN3-GS^{yellow}) - $\log_2$ average LFQ intensity (group2, WT); calculated from the Proteingroups file (Supplemental Data S2E) with the Perseus software (as described in the methods section, briefly: LFQ intensities were logged; missing values were imputated from a normal distribution around the detection limit of the mass spectrometer; average $\log_2$ LFQ intensity was calculated for each identification in both groups: AN3 (green columns) and WT (Col-0, yellow columns); for each identification, the difference in average $\log_2$ LFQ intensity between the 2 groups is calculated and shown here.
[2]-LogP: Test statistic: $-\log_{10}$ p-value, taking into account FDR=0.01 and S0=0.9

The usability of the 35S promoter in seedling TAP experiments is already shown in Chapter 3. There, co-purification of the AN3-associated SWI/SNF complex through TAP-MS with the GS^{yellow} was evaluated. Our data showed that both *Arabidopsis* cell cultures and *Arabidopsis* seedlings expressing 35S:AN3-GS^{yellow} provided a high overlap with previously reported data obtained through TAP-MS using the GS^{rhino} tag. Also, testing the functionality of the construct through

overexpression of the recombinant AN3 in a WT background resulted in the expected enlarged leaf phenotype (Chapter 3). Using these results as a benchmark, a pull-down experiment was performed on the same biological material to see how TAP relates to pull-down when strong constitutive promoters are driving the expression of the bait protein.

A long list of possible interactors was retained using standard filtering methods (Table S3 and Supplemental Data S2). Strikingly, only six of the known SWI/SNF interactors (ARP7, BRM, CHLI1, G2482-1, SWI3C and SWP73B) were found to be significantly present in the transgenic samples compared to WT (see Table 3 and Figure S2), whereas with the TAP experiment almost all known subunits of the SWI/SNF complex were identified, as well as some TFs (WOX13 and GRF3). Moreover, AN3 itself was not retained in the pull-down because it was only found with a single peptide in the dataset. Besides the six interactors, which are subunits of the reported SWI/SNF core complex, a lot of significantly enriched proteins (compared to WT) were obtained (Table S3) that were not previously found in AN3 purification experiments. Most likely they are false positive interactors due the overexpression of the bait protein and the lower purity level in pull-down experiments as compared to TAP.

COMING CLOSER: AP-BASED DIFFERENTIAL COMPLEX ANALYSIS IN *ZEA MAYS*

Besides the implementation of quantitative AP-MS as a means to help the differentiation between *bona fide* interaction partners of the bait and false positives, it can also be used to see how interactions change in different experimental contexts. Chapter 3 showed how TAP in combination with MaxLFQ is able to capture such dynamics by analysing the changes in interaction partners of the transcriptional co-activator ZmAN3 along the different developmental stages (*i.e.*, division versus expansion zone) of a growing *Zea mays* leaf. Here, the experiment was repeated by performing pull-down on the same line, and the results (see Table 4 and Supplemental Data S3) were filtered according to standard practices using WT B104 plants as control (see Chapter 3; also see pull-down in *Arabidopsis* above). Strikingly, both methods delivered similar data regarding the non-dynamic interaction of the SWI/SNF subunits. To allow identification of more interaction partners in the pull-down on the expansion zone, FDR values had to be increased from 0.01 to 0.05 (see Supplemental Data S3). In total, six of the SWI/SNF subunits picked up by TAP-MS in the division zone were absent in the pull-down results, and no GRFs could be qualitatively detected in the expansion zone pull-down. Quantitative analysis on the LFQ values shows higher abundance of the GRFs in the division zone, thus confirming the dynamics that were denoted previously with TAP-MS (Nelissen *et al.*, 2015), and Chapter 3). Besides the known interactors, pull-down also gives a list of 33 other 'enriched' proteins that could not be filtered out based on the LFQ differences (Table S4). However, nine of them are present in the extended 'background list' based on orthologues of known co-purifying proteins identified in *Arabidopsis* TAPs.

Table 4. Interaction partners of ZmAN3 identified through pull-down on dividing or expanding leaf cells

ZmAN3		Division Zone		Expansion Zone		Previously Reported
Protein ID	Protein Name	* Difference[1]	- Log P[2]	** Difference[1]	-Log P[2]	
Core Subunits						
GRMZM2G180246	AN3	13.91	6.21	13.14	4.95	x
GRMZM2G015384	ARP4	12.17	5.79	10.99	3.35	x
GRMZM2G015861	ARP7	11.35	4.14	9.27	3.57	x
GRMZM2G018955	BCL7	6.47	4.62	4.44	2.76	x
GRMZM2G020548	BRD1	5.03	2.38			x
GRMZM2G044044	BRD1	5.99	3.73			x
GRMZM2G312501	BRD1	5.18	3.51			x
GRMZM2G336962	BRD1	5.76	3.46			x
GRMZM2G163849	BRM	15.61	3.76	13.29	5.00	x
GRMZM5G816791	dentin sialophosphoprotein-related			3.85	2.96	x
GRMZM2G103079	dentin sialophosphoprotein-related	7.36	5.35			x
GRMZM2G473310	G2484-1	7.32	2.19	8.52	2.49	x
GRMZM2G033478	G2484-1			4.30	2.70	x
AC198518.3_FGP003	GLTSCR	9.78	4.96	9.08	5.17	x
GRMZM2G476652	LFR	6.66	3.11	6.54	3.42	x
GRMZM2G139760	SWI3C	8.32	4.06			x
GRMZM2G340756	SWI3C	7.75	3.61	8.68	3.80	x
GRMZM2G119261	SWI3D	7.09	3.88	5.34	3.32	x
GRMZM2G052416	SWP73B	9.39	4.53	8.79	4.59	x
GRMZM2G387890	SYD	10.94	4.75	11.05	3.97	x
GRMZM2G467799	SYD	4.96	3.06			x
GRFs						
GRMZM2G034876	GRF1	6.87	2.43			x
GRMZM2G096709	GRF10/GRF4	6.70	4.30	5.54	4.24	x
GRMZM2G067743	GRF11	8.72	3.70			x
GRMZM2G119359	GRF12	7.13	4.41			x
GRMZM2G178261	GRF15	6.37	3.01			x
GRMZM2G124566	GRF17/GRF18	8.76	6.01			x
GRMZM2G105335	GRF3	7.79	5.00			x
GRMZM2G129147	GRF5	6.28	3.81			x
GRMZM5G850129	GRF7	6.57	5.30			x

* Compared to WT, FDR = 0.01

** Compared to WT, FDR = 0.05

[1]Difference: $\log_2$ (LFQ ratio (ZmAN3-GS^{yellow}/WT)) = $\log_2$ average LFQ intensity (group1, ZmAN3-GS^{yellow}) - $\log_2$ average LFQ intensity (group2, WT); calculated from the Proteingroups file (Supplemental Data S3A) with the Perseus software (as described in the methods section, briefly: LFQ intensities were logged; missing values were imputated from a normal distribution around the detection limit of the mass spectrometer; average $\log_2$ LFQ intensity was calculated for each identification in both groups: Group 1: ZmAN3 division zone (green columns) or ZmAN3 expansion zone (red columns), and WT (group 2) (yellow columns); for each identification, the difference in average $\log_2$ LFQ intensity between the 2 groups is calculated and shown here.

[2]-LogP: Test statistic: $-\log_{10}$ p-value, taking into account FDR= see * or ** and So=0.1.

DISCUSSION

AP-MS is one of the most commonly applied approaches for studying protein complexes in the field of functional genomics. By fusing an affinity tag to the protein of interest, the technique allows the purification of associated protein complexes and identifies the subunits in a high-throughput manner. However, besides the complex of interest, many false positive interactors and commonly co-purifying proteins are also picked up. At present, there are two strategies to avoid

false positive interactors: either filtering them by implementing the quantitative data that are gathered during MS to identify true interactors from false ones, or reducing the number of false positives by adding a second round of purification to reduce the amount of co-purifying proteins. To gain an insight into the differences gathered with both approaches, we conducted a set of pull-down and TAP experiments in several plant backgrounds. By tagging a handful of baits with the GS^{yellow} tag, both purification methods could be performed on the same transgenic material and allowed, as such, a direct comparison of the results.

With the results obtained for the TAPs on SAMBA-GS^{yellow} in *Arabidopsis* cell suspension cultures, data from previous experiments on SAMBA-GS^{TEV} in both cell cultures (Van Leene *et al.*, 2010) and seedlings (Eloy *et al.*, 2012) could be confirmed. By picking up the same APC/C core complex (with the exception of APC3a), and some co-purified regulators (CCS52A2, UVI4 and UVI4-like), more known interactors were identified with TAP than with pull-down. Moreover, given the complete absence of the regulatory subunits of the complex in the pull-down eluate, our data thus confirm the strength of TAP in picking up substoichiometric interactions. The reason for this can be found in the lower purity of pull-down eluates, which interfere with the detection of such substoichiometric interactions.

In *Arabidopsis* seedlings, on the other hand, the comparison showed different results for the two baits. For the seedlings expressing 35S:AN3-GS^{yellow}, TAP provided more components of the SWI/SNF complex that were previously reported to associate to AN3 (Vercruyssen *et al.*, 2014) than pull-down. The TAP results also confirmed the reported interactions of AN3 with GRF3 and WOX13, which were absent in the pull-down data. Remarkably, only a single peptide for AN3 was detected in the pull-down samples. It could be that in these experiments, the large amount of co-purifying proteins for which no interaction with AN3 has been previously reported interfere with the detection of both AN3 and its true interactors. A similar yet different result was seen for the TAP versus pull-down experiments on seedlings expressing AUR1-GS^{yellow}. From the TAP results in cell cultures, we established that the predicted TPX2-like proteins indeed bind to AUR1 (further validated through Y2H for TPXL2), and that TAP can be used to identify associating protein complexes of AUR1. Remarkably, the TAP on seedlings could only confirm the interaction with TPXL3, whereas TPX2 or TPX-like proteins were not picked up in the pull-down experiments. The latter did however pick up AUR2, which was shown to associate with AUR1 through TAP on cell cultures that expressed an N-terminally tagged TAPi-AUR1 (an optimised and codon-adapted version of the TAP tag for plants (Rohila *et al.*, 2004)), and through Y2H (Yi and Li, 2012). For the other two proteins that were retained with pull-down, no previous interactions were reported, which might indicate that they are either new interactors, or that they are unspecific interactors or background proteins.

The third comparison was carried out on ZmAN3 in two developmentally different stages (*i.e.*, division and expansion zones) of the leaf growth zone. Here, the pull-down on the expansion zone clearly performed less well than TAP-MS because the FDR had to be increased from 0.01 to 0.05 to allow the identification of

more interaction partners (see Supplemental Data 3), and still six of the SWI/SNF subunits picked up by TAP-MS in the division zone were absent in the pull-down results. Moreover, the qualitative absence of almost all GRFs in the expansion zone identified with pull-down further indicates the lesser performance of the pull-down. We cannot exclude that the difference in input, which was almost half for the pull-down compared to TAP experiments, is an explanation for the difference in results. However, one could speculate that the percentage of complexes retained with pull-down is much higher, as the captured bait is subjected to less stringent wash steps than in TAP. Still, the quantitative analysis on the LFQ values shows z higher abundance of the GRFs in the division zone, thus confirming the dynamics that were denoted previously with TAP-MS (Nelissen *et al.*, 2015; and Chapter 3). Besides the known interactors, pull-down also provided a list of 33 other 'enriched' proteins. However, as nine of them are present in the extended 'background list' based on orthologues of known co-purifying proteins identified in *Arabidopsis* TAPs, these might not represent true interactions.

In addition to the comparison of pull-down versus TAP, two different promoters (35S and endogenous pAUR1) and genomic backgrounds in *Arabidopsis* (WT and *aur1-2aur2-2* double mutant) were also tested with AN3 and AUR1. The obtained data show that even when the selected endogenous promoter can complement the mutant phenotype, as was demonstrated here for AUR1, its expression can be too low to allow identification of all true interaction partners. Nevertheless, some previously reported or predicted interactions could be confirmed, including the heterodimerisation of AUR1 with AUR2, as well as the interaction with TPXL3. Compared to the cell culture results, however, the seedling experiments of AUR1 clearly had lower signal-to-noise ratios which are caused by the usage of the endogenous regulatory sequences. Although seedlings are reported to have a high ratio of dividing cells (Vercruyssen *et al.*, 2014), this proportion is possibly still too low for the detection of interaction partners during cell division. Use of a strong constitutive promoter, such as the 35S, which comes at the cost of losing the developmental context and timing of the interactions, was able to provide more interactions for AN3. This indicates that using an overexpression promoter in a WT background might still be beneficial for the identification of protein complexes.

CONCLUSION

A PROTEIN ANALYSIST'S GUIDE TO AFFINITY PURIFICATION

Choosing between pull-down and TAP to identify a protein's interaction partners via MS is a miscellaneous issue. A first point of consideration is the amount of biological material that is needed for the experiment. TAP generally requires three-to-ten times more input material than AP (Van Leene *et al.*, 2014). In plant species, such as *Zea mays*, where upscaling is rather time and resources consuming, decreasing the amount of biomass is thus a great advantage. Here, zooming in on specific tissues of the developing leaf led to a reduction of the unspecific and sticky proteins in non-target tissues, which is in favour of pull-downs that normally suffer

from high background levels. In cases where input materials are unlimited, such as the *Arabidopsis* cell cultures, this advantage is less profound. Besides the characteristics of the biological material, the obtained data also show how the performance of the technique differs for each bait. As most regulatory interactions in the cell are transient in nature, using pull-down for the detection of short binding partners might be a better choice, although this was not observed in our experiments. In case the bait incorporates into a stable complex, as was shown here for AN3 and SAMBA, reducing the amount of background by a second purification round (*i.e.*, TAP) clearly provides better results in the detection of substoichiometric interactions. Hence, when there is no problem concerning the generation of biological material, we advise to perform a TAP experiment, as our data showed that TAP with the GS$^{yellow/green}$ outperformed standard GFP pull-down.

Our data also suggest that great care has to be given to the selection of the promoter in AP-MS experiments. Although endogenous promoters provide a more biologically specific view on the interaction patterns of the bait and are becoming more popular in biological studies, the time frame or developmental stages of expression are sometimes too specific or too low to pick up interactors. Using the strong constitutive 35S promoter, on the other hand, provides sufficient bait accumulation for the detection of PPIs, and is a necessity if a WT background is used. However, it can lead to false positives as the bait is expressed in cells where it is absent under normal conditions.

In conclusion, rather than limiting the choice between pull-down or TAP, using tags, such as the GS^{yellow} TAP tag that can be used for both approaches, is a better option. Another idea is also to provide a pull-down protocol for other TAP tags, such as the GS^{rhino} TAP tag, a strategy that is currently being evaluated in our group and shows promising results. Both methods allow the selection of the purification scheme based on the experimental setting and the available biomass rather than on the implemented tag. This circumvents the need of generating different transgenic lines, allowing direct comparison of all obtained data. As such, they could provide the research community with a flexible tool to help the functional analysis of PPIs in plants.

MATERIALS AND METHODS

CONSTRUCTION OF ENTRY CLONES

The Gateway (Invitrogen) recombination-mediated construction of the pENL1L2 entry clones containing SAMBA (Eloy *et al.*, 2012), AUR1 and AUR2 (Van Damme *et al.*, 2012), AN3 (Vercruyssen *et al.*, 2014) or ZmAN3 (Nelissen *et al.*, 2015) are described elsewhere. Also the production of the pENL4R1_proAUR1 (Van Damme *et al.*, 2012), pENL4R1_proUBIL (Karimi *et al.*, 2007) and the pENL2R3_GS$^{yellow/green}$ (C-terminal) and pENL4R1_NGS$^{yellow/green}$ (N-terminal)(see Chapter 3, Materials and Methods) is reported previously. Both TPXL2 and TPXL3 were picked up as coding sequence from *Arabidopsis* cDNA with gene-specific primers flanked by the necessary *att* sites. For Y2H, the pENL1L2_AUR1,

pENL1L2_AUR2, pENL1L2_TPXL2 and pENL1L2_TPXL3 were cloned into the pDEST22 (contains the GAL4-AD) and pDEST32 (contains the GAL4-DBD) through a Gateway LR reaction.

CONSTRUCTION OF EXPRESSION VECTORS

35S:ORF tag or 35S:tag-ORF constructs were generated as described (Van Leene *et al.*, 2007). pAUR1:AUR1-GS^{yellow} was produced through a single Gateway LR reaction with the corresponding entry vectors and the pB7m34GW destination vector. The construction of p35S:AN3-GS^{yellow} and pUBIL:ZmAN3-GS^{yellow} was described in a previous chapter (Chapter 3, Materials and Methods).

PLANT MATERIAL AND GROWTH CONDITIONS

PSB-D *Arabidopsis thaliana* cell suspension cultures were transformed with 35S:ORF tag constructs, selected without callus screening and subcultured as described by (Van Leene *et al.*, 2007). Transformation of *Arabidopsis* plants was carried out through floral dip. 35S:AN3-GS^{yellow} was transformed into *Arabidopsis thaliana (Col-0*) plants and selected as described in Chapter 3 (Materials and Methods). pAUR1:AUR1-GS^{yellow} was transformed into the *aur1-2aur2-2* mutant background via floral dip, and transgenic offspring was selected on ½ MS medium supplemented with 1% (w/v) sucrose and Basta. For complex analysis of 35S:AN3-GS^{yellow} using pull-down or TAP, *Arabidopsis* seedlings were grown as described by (Van Leene *et al.*, 2015). For complex analysis on pAUR1:AUR1-GS^{yellow} through TAP, *Arabidopsis* seedlings were grown for 4 days as described (Van Leene *et al.*, 2015), whereas for pull-down, seedlings were grown for 3 days on plates containing ½MS medium supplemented with 1% (w/v) sucrose. For expression analysis using immunoprecipitation, pAUR1:AUR1-GS^{yello}-expressing seedlings were grown in liquid ½MS medium for 2, 3 or 4 days as described in Chapter 3 (Materials and Methods). *Zea mays* plants containing pUBIL:ZmAN3-GS^{yellow} were grown and harvested as described in Chapter 3 (Materials and Methods).

LATERAL ROOT DENSITY MEASUREMENTS

Lateral root densities were measured as described (Van Damme *et al.*, 2012). Briefly, seedlings were grown for 12 to 13 days on vertical plates under continuous light. Late germinating seeds were scored 2 days after vernalisation and excluded from further analysis. Emerged lateral roots were counted using a Leica S4E stereomicroscope, and main roots were measured from plate scans obtained with Epson perfection V700 photo scanner at 300 dpi using ImageJ(Fiji).

PROTEIN PURIFICATION AND EXPRESSION ANALYSIS

Expression analysis through Western blot of purified proteins was carried out as described (Van Leene *et al.*, 2007; Gadeyne *et al.*, 2014, Nelissen *et al.*, 2013). Primary detection of GFP/cYFP (GS$^{green/yellow}$)-tagged proteins was obtained with anti-GFP (*ab290*, Abcam©), whereas GS^{rhino} was detected using Peroxidase Anti-Peroxidase (Sigma). For expression analysis of pAUR1:AUR1-GS^{yellow} in the *aur1-*

2aur2-2 mutant background, an immunoprecipitation was carried out as follows: 1-2 g of 2-, 3- or 4-day-old seedlings were harvested, frozen in liquid nitrogen and crushed manually to powder with mortar and pestle. Two volumes of BHB+ extraction buffer (see (Van Leene *et al.*, 2014) with 1/1000 benzonase were added to each sample and incubated for 30 minutes at 4°C using mild rotation. Next, the samples were centrifuged for 20 minutes at maximum speed, 4°C, transferred to a new tube and subjected to a second round of centrifugation at maximum speed, 20 minutes, 4°C. Protein concentrations were measured on the obtained supernatant using Bradford assays, and equal protein inputs were given (4.1 mg) for all samples. The appropriate amount of supernatant was then incubated with 20 µL anti-GFP agarose beads (GFP-Trap®_A, Chromotek) for 1 h at 4°C using mild rotation. Beads were recovered on mobicols and washed with 150 KV TEV buffer (see (Van Leene *et al.*, 2015)). Bound fractions were eluted by boiling for 10 min at 95°C in 20 µL of sample buffer (see (Van Leene *et al,* 2015)), and loaded on SDS-PAGE for Western blot analysis.

TANDEM AFFINITY PURIFICATIONS

TAP experiments on transformed cell suspension cultures were carried out as described by (Van Leene *et al.*, 2007) with 100 mg full protein input. TAP experiments on *Arabidopsis* seedlings were performed using a published (Van Leene *et al.*, 2015). For the AUR1-GS^{yellow} and AN3-GS^{yellow} seedling TAPs respectively, 100 g and 50 g seedlings were used as input material.

For the GS^{yellow} constructs we used GFP-trap beads (GFP-Trap®_A, Chromotek) for the first purification round, followed by a second round of purification with streptavidin beads. Primary elution was done with the human Rhino virus 3C protease.

PULL-DOWNS

The pull-down experiments on pUBIL:ZmAN3-GS^{yellow} were carried out on zone-specific samples coming from 30 plants per repeat, with equal full protein inputs for all samples (8 mg), using the described protocol (Wendrich *et al.*, 2017). For pull-downs on *Arabidopsis* seedlings expressing pAUR1:AUR1-GS^{yellow}, the same (Wendrich *et al.*, 2017) protocol was used, starting from 2,3 g seedlings. For the pull-downs on p35S:AN3-GS^{yellow}-expressing *Arabidopsis* seedlings, 2.5 gram biological material was subjected to an adapted version of the (Wendrich *et al.*, 2017) protocol as follows: ground material was resuspended in 1.5x (w/v) BHB+ buffer supplemented with 1.5x µl benzonase and incubated for 30 min at 4°C. Samples were mixed 3 times for 30 s at 18,000 rpm using Ultra-TURRAX miniprobes. Next, samples were incubated for 30 min at 4°C using an end-over-end rotor, and spinned down twice at 14,000 rpm for 20 min in an Eppendorf centrifuge, with transfer to new Eppendorf tubes in-between centrifugations. The supernatant was filtered using a 0.45-µm filter, and 50 µL anti-GFP-beads (µMacs) were added to the extract. Samples were then incubated for 1 h at 4°C in an end-over-end shaker. Next, they were loaded on prewashed µMacs columns and washed 4times with

200 µL BHB+ buffer (see (Van Leene *et al.*, 2015) and two times with 500 µl NH_4HCO_3 buffer Finally, bound proteins were eluted twice with 50 µL 95°C hot NH_4HCO_3. Per sample, 4 µL trypsin/lysine mix was added and proteins were digested on beads for 4 h at 37°C with agitation (800 rpm on a thermomixer). The protein eluates were then separated from the beads using the µMacs columns, supplemented with 2 µL trypsin/lysine mix and digested overnight at 37°C with agitation (800 rpm on a thermomixer). Samples were then frozen quickly in liquid nitrogen and dried in a Speedvac. MS analysis was carried out as described (Van Leene *et al.*, 2014).

IDENTIFICATION OF INTERACTION PARTNERS FROM PULL-DOWNS

The obtained LC-MS/MS data were analysed using the Andromeda algorithm for label-free quantification (LFQ) from MaxQuant 1.2.0.5 according to the protocol (Wendrich *et al.*, 2017). All settings and implemented databases are described in Chapter 3. Shortly, LFQ intensities were normalised using the quantitative information obtained for all protein groups that were identified with at least two peptides. Replicates were grouped together. Peptide searches were carried out using matching between runs to identify all peptides in all MS runs through their retention time, measured *m/z* value and isotopic distributions. Next, *t*-tests between two groups on the obtained LFQ intensities were carried out using Perseus, resulting in a p-value (-log *t*-test p-value) and a difference (ratio of the group's average LFQ values; *t*-test difference). For statistical analysis, FDR was set at 1% to restrict the results to highly reliable interactions. So values were set at 1 (Wendrich *et al.*, 2017).

MICROSCOPY

Confocal microscopy of dividing cells was carried out on hypocotyl cells as described (VanDamme *et al.*,2012) using a FluoView1000 inverted confocal microscope (Olympus) equipped with a water-corrected ×60 objective (numerical aperture of 1.2). Laser excitation was carried out at 515 nm for cYFP detection, and its spectrum was exclusively detected between 530-560 nm. Cell walls were stained before image acquisition with propidium iodide, which was visualised using 514-nm laser excitation and 566- to 649-nm spectral detection.

REFERENCES

Coutuer S. 2008. Cytoskeleton-associated protein complexes during plant cell division.

Dedecker M, Van Leene J, De Jaeger G. 2015. Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry. Current Opinion in Plant Biology **24**, 1–9.

Dedecker M, Van Leene J, De Winne N, *et al*. 2016. Transferring an optimized TAP-toolbox for the isolation of protein complexes to a portfolio of rice tissues. Plant Molecular Biology **91**, 341–354.

Eloy NB, de Freitas Lima M, Ferreira PCG, Inzé D. 2015. The Role of the Anaphase-Promoting Complex/Cyclosome in Plant Growth. Critical Reviews in Plant Sciences **34**, 487–505.

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 GT, Remaut H, Van Montagu MC, De Jaeger G, Ferreira PC, Inzé 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.

Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Höfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415**, 141–147.

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. The Plant Journal **88**, 476–489.

Karimi M, Bleys A, Vanderhaeghen R, Hilson P. 2007. Building blocks for plant gene assembly. Plant Physiology **145**, 1183–1191.

Keilhauer EC, Hein MY, Mann M. 2015. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). Molecular & Cellular Proteomics **14**, 120–135.

Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY, Halim VA, Bagshaw RD, Hubner NC, Al-Hakim A, Bouchard A, Faubert D, Fermin D, Dunham WH, Goudreault M, Lin ZY, Badillo BG, Pawson T, Durocher D, Coulombe B, Aebersold R, Superti-Furga G, Colinge J, Heck AJ, Choi H, Gstaiger M, Mohammed S, Cristea IM, Bennett KL, Washburn MP, Raught B, Ewing RM, Gingras AC, Nesvizhskii AI. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nature Methods **10**, 730–736.

Meyer K, Selbach M. 2015. Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. Frontiers in genetics **6**, 237.

Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. 2011. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Molecular & Cellular Proteomics **10**, M111.011015

Mittler G, Butter F, Mann M. 2008. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Research **19**, 284–293.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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..

Pardo M, Choudhary JS. 2012. Assignment of Protein Interactions from Affinity Purification/Mass Spectrometry Data. Journal of Proteome Research **11**, 1462–1474.

Rohila JS, Chen M, Cerny R, Fromm ME. 2004. Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. The Plant Journal **38**, 172–181.

Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, Zhu JK, Ronald P, Fromm ME. 2006. Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. The Plant Journal **46**, 1–13.

Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. 2008. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Molecular & Cellular Proteomics **7**, 282–289.

Tomaštíková E, Demidov D, Jeřábková H, Binarova P, Houben A, Doležel J, Petrovská B. 2015. TPX2 Protein of *Arabidopsis* Activates Aurora Kinase 1, But Not Aurora Kinase 3 In Vitro. Plant Molecular Biology Reporter **33**, 1988–1995.

Van Damme D, De Rybel B, Gudesblat G, Demidov D, Grunewald W, De Smet I, Houben A, Beeckman T, Russinova E. 2012. *Arabidopsis*α Aurora Kinases Function in Formative Cell Division Plane Orientation. Plant Cell **23**, 4013–4024.

Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L. 2011. A kaleidoscopic view of the

Arabidopsis core cell cycle interactome. Trends in Plant Science **16**, 141–150.

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. Nature Protocols **10**, 169–187.

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, Inzé D, Witters E, De Jaeger G. 2010. Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis* thaliana. Molecular Systems Biology **6**, 1–12.

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, Inzé D, Witters E, De Jaeger G. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis* thaliana. Molecular & Cellular Proteomics **6**, 1226–1238.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L, Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. **26**, 210–229.

Wendrich JR, Boeren S, Möller BK, Weijers D, De Rybel B. 2017. In Vivo Identification of Plant Protein Complexes Using IP-MS/MS. Methods Mol Biol **1497**:147-158.

Yi K, Li R. 2012. Actin cytoskeleton in cell polarity and asymmetric division during mouse oocyte maturation (DN Robinson, WM Bement, MK Balasubramanian, and JW Sanger, Eds.). Cytoskeleton **69**, 727–737.

SUPPLEMENTAL FIGURES AND TABLES

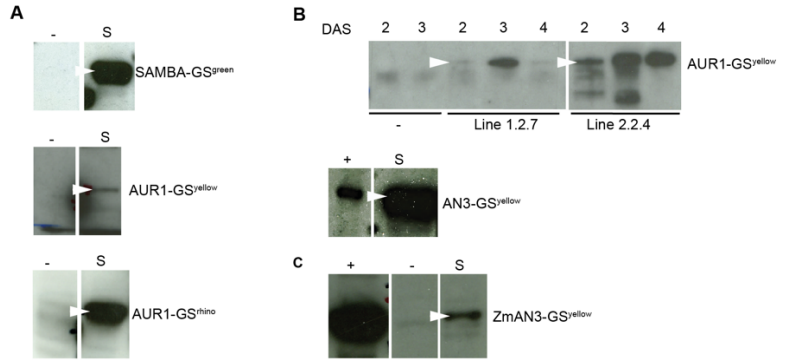


Figure S1. Western blot analysis of protein expression of all lines used for pull-down or TAP. **(A)** *Arabidopsis* cell cultures transformed with SAMBA-GS^{green}, AUR1-GS^{yellow} or AUR1-GS^{rhino}, all expressed from the 35S promoter. -; Mock (WT) culture. **(B)** *Arabidopsis* seedlings. **Top**: *aur1-2 aur2-2* double mutant seedlings transformed with pAUR1:AUR1-GS^{yellow} were subjected to pull-down analysis to assess the time-point (Days After Sowing) at which AUR1-GS^{yellow} expression is the highest. -; WT (Col-0) seedlings; **Bottom**: Col-0 seedlings expressing p35S:AN3-GS^{yellow}. **(C)** Proteins extracted from leaf 4 of *Zea mays* plants transformed with pUBIL:ZmAN3-GS ^{yellow}; -; WT (B104) line; white arrows indicate the fusion protein

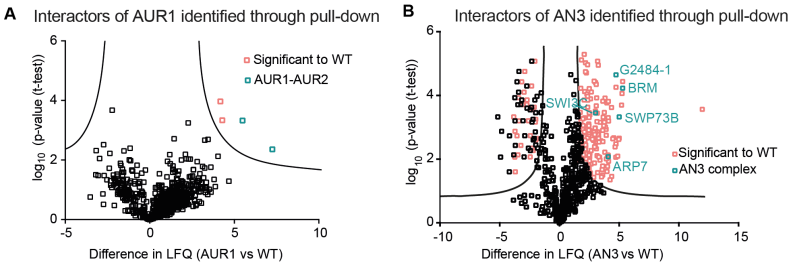


Figure S2. Relative Intensity-Based Label-Free Quantification of interaction partners identified by pull-down in seedlings of *Arabidopsis* as compared to WT. Significantly enriched interactors were identified through permutation-based FDR-corrected *t*-test (threshold FDR = 0.01 and S0 = 0.9; Supplemental Table S2), which is indicated on the plot by the curve(s). The difference in average $\log_2$ LFQ intensities between both samples is plotted versus their significance ($-\log_{10}$(p-value)). **(A)** The known interactor AUR2, indicated in green, was retained when performing pull-down on 3-day-old complemented *aur1-2aur2-2* double mutant seedlings expressing AUR1-GS^{yellow} from its endogenous promoter. Two additionally significantly enriched proteins that were not previously reported to interact with AUR1 are indicated in red. **(B)** Pull-down on 6-day-old Col-0 seedlings transformed with 35S:AN3-GS^{yellow} provided five subunits (indicated in green) of the SWI/SNF complex. Other significantly enriched proteins are depicted in red. AN3 itself was only detected with a single peptide and is therefore not presented in the plot.

Supplemental Table S1. List of co-purifying proteins picked up through pull-down on SAMBA in *Arabidopsis* plant cell cultures (Van Leene *et al.*, 2015) that are considered background

SAMBA		AP		TAP
ProteinID	Protein name	Difference	LogP	back-ground
AT5G25460	F18G18_200	13.84	4.46	x
AT5G14040	phosphate transporter 3;1	11.11	3.8	x
AT3G08580	ADP/ATP carrier 1	9.65	7.35	x
AT2G20800	NAD(P)H dehydrogenase B4	9.65	4.23	x
AT3G55430	O-Glycosyl Hydrolases Family 17 Protein	9.49	3.46	
AT3G08030	Protein of unknown function, DUF642	9.22	1.79	x
AT1G79530	glyceraldehyde-3-phosphate dehydrogenase of plastid 1	8.92	3.97	x
AT5G12470	Protein of unknown function (DUF3411)	8.34	3.98	x
AT1G22650	Plant Neutral Invertase Family Protein	7.99	4.53	
AT4G01100	adenine nucleotide transporter 1	7.92	4.63	x
AT4G30010		7.83	3.66	
AT5G13490	ADP/ATP carrier 2	7.77	3.96	x
AT5G13430	Ubiquinol-Cytochrome C Reductase Iron-Sulfur Subunit	7.67	5.26	
AT3G53520	UDP-Glucuronic Acid Decarboxylase 1	7.62	3.93	
AT3G10410	Serine Carboxypeptidase-Like 49	7.53	2.43	
AT2G41800	Protein of unknown function, DUF642	7.39	2.12	x
AT3G62830	NAD(P)-binding Rossmann-fold superfamily protein	7.28	1.41	x
AT3G52730	Ubiquinol-Cytochrome C Reductase UQCRX/QCR9-Like Family Protein	7.15	2.58	
AT2G35900		7.15	4.72	
ATMG01190	ATP synthase subunit 1	6.84	4.5	x
AT5G59850	Ribosomal protein S8 family protein	6.77	2.88	x
AT1G28110	Serine Carboxypeptidase-Like 45	6.73	3.56	
AT5G09810	actin 7	6.58	2.45	x
AT3G44110	DNAJ homologue 3	6.53	3.61	x
AT3G52990	Pyruvate kinase family protein	6.44	4.52	x
AT3G62530	ARM repeat superfamily protein	6.39	2.57	x
AT1G74000	Strictosidine Synthase 3	6.37	4.88	
AT5G66920	SKU5 Similar 17	6.36	4.03	
AT4G37910	mitochondrial heat shock protein 70-1	6.31	5.63	x
AT5G08690	ATP synthase alpha/beta family protein	6.27	2.64	x
AT5G64260	EXORDIUM like 2	6.06	4	x
AT3G53870	Ribosomal protein S3 family protein	6.04	5	x
AT3G61820	Eukaryotic Aspartyl Protease Family Protein	6	3.74	
AT3G27400	Pectin lyase-like superfamily protein	5.94	3.78	x
AT4G34200	D-3-phosphoglycerate dehydrogenase	5.93	3.69	x
AT3G02530	TCP-1/cpn60 chaperonin family protein	5.92	4.42	x
AT5G67630	RUVB2-like	5.84	4.58	x
AT3G04120	glyceraldehyde-3-phosphate dehydrogenase C subunit 1	5.8	5.03	x
AT3G18780	actin 2	5.76	3.85	x
AT4G30190	H(+)-ATPase 2	5.65	3.3	x
AT1G33140	Ribosomal protein L6 family	5.35	4.48	x
AT1G71380	cellulase 3	5.28	3.15	x
AT1G06950	translocon at the inner envelope membrane of chloroplasts 110	5.27	4.17	x
AT1G01300	Eukaryotic aspartyl protease family protein	5.23	4.11	x
AT5G42150	Glutathione S-Transferase Family Protein	5.17	4.04	
AT3G20050	T-complex protein 1 alpha subunit	5.05	5.3	x
AT1G72750	Translocase Inner Membrane Subunit 23-2	5.04	4.05	
AT5G16070	TCP-1/cpn60 chaperonin family protein	5.01	2.29	x
AT1G80240	DUF642 L-Gall Responsive Gene 1	4.91	1.47	
AT2G20940	Protein Of Unknown Function (DUF1279)	4.74	2.91	
AT5G05010	clathrin adaptor complexes medium subunit family protein	4.67	5.13	x
AT5G21150	Argonaute Family Protein	4.62	3.21	
AT3G62290	ADP-ribosylation factor A1E	4.6	3.67	x
AT1G79050	recA DNA recombination family protein	4.54	3.09	x

AT5G50850	Transketolase family protein	4.51	2.54	x
AT4G38400	Expansin-Like A2	4.03	4.04	
AT4G18800	RAB GTPase homolog A1D	4	1.67	x
AT3G46060	RAB GTPase homolog 8A	3.95	2.12	x
AT5G28540	heat shock protein 70 (Hsp 70) family protein	3.78	2.19	x
AT3G58510	DEA(D/H)-box RNA helicase family protein	2.96	0.68	x
AT3G11940	ribosomal protein 5A	2.66	3.13	x
AT2G41840	Ribosomal protein S5 family protein	2.55	1.65	x
AT5G60390	GTP binding Elongation factor Tu family protein	2.09	0.9	x
AT1G41830	SKU5-Similar 6	1.6	1.56	
AT2G36200	P-Loop Containing Nucleoside Triphosphate Hydrolases Superfamily Protein	1.26	0.45	
AT4G09800	S18 ribosomal protein	0.99	0.6	x

Supplemental Table S2. List of AN3 interactors considered background as found with pull-down on *Arabidopsis* seedlings.

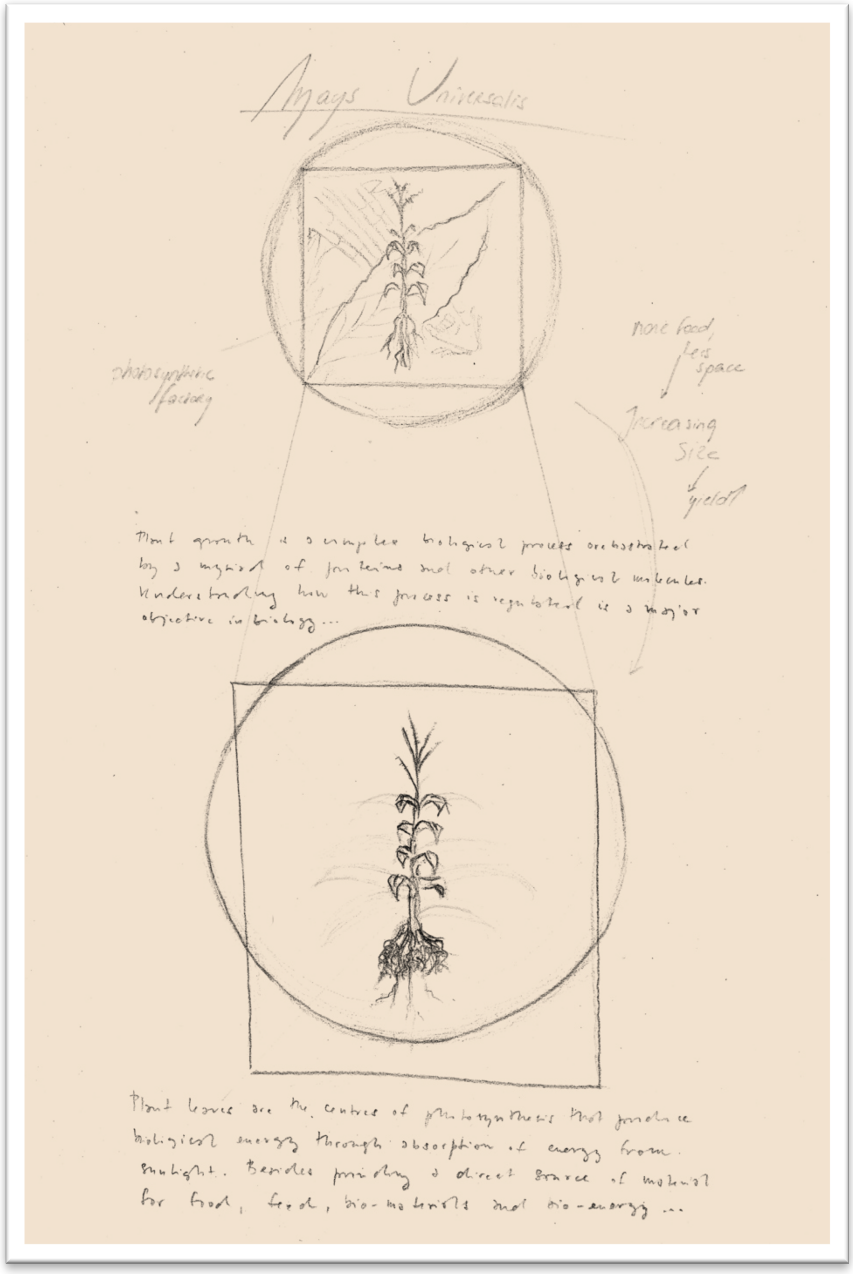
AN3		**AP**		**TAP back-ground**
ProteinID	**Protein name**	**Difference**	**LogP**	
AT3G17390	MTO3	11.96	4.4	x
AT3G48870	HSP93-III	5.25	3.92	x
AT4G11420	EIF3A	5.2	2.59	x
AT3G27850	RPL12-C	5.01	3.33	x
AT1G74470	Pyridine nucleotide-disulphide oxidoreductase family protein	4.97	2.57	x
AT5G28840	GME	4.88	3.44	
AT5G60390	GTP-binding Elongation factor Tu family protein	4.74	4.65	x
AT5G20000	AAA-type ATPase family protein	4.73	2.14	x
AT5G46580	pentatricopeptide(PPR)repeat-containing protein	4.73	2.91	x
AT1G09310	unknown protein	4.66	4.22	
AT5G26742	emb1138	4.55	5.12	x
AT3G54210	Ribosomal protein L17 family protein	4.47	2.06	x
AT5G42080	DL1	4.38	2.06	x
AT3G59020	ARM repeat super family protein	4.36	1.44	x
AT5G41670	6-phosphogluconate dehydrogenase family protein	4.24	3.62	x
AT1G79930	HSP91	4.12	1.61	x
AT1G62380	ACO2	4.11	4.75	x
AT2G35840	Sucrose-6F-phosphate phosphohydrolase family protein	4.1	4.45	
AT4G29040	RPT2a	4.09	2.93	x
AT3G63460	Transducin family protein/WD-40 repeat family protein	4.09	2.08	x
AT2G36250	FTSZ2-1	4.04	4.96	x
AT1G20950	Phosphofructokinase family protein	4.03	2.48	
AT4G39520	GTP-binding protein-related	3.98	3.16	x
AT3G55610	P5CS2	3.98	2.11	x
AT2G05830	NagB/RpiA/CoA transferase-like super family protein	3.94	4.04	x
AT1G22360	AtUGT85A2	3.92	3.77	
AT2G38040	CAC3	3.91	1.96	
AT5G51110	Transcriptional co-activator/pterin dehydratase	3.85	2.09	x
AT2G21390	Coatomer	3.82	2.17	x
AT1G29930	CAB1	3.8	4.35	x
AT3G02360	6-phosphogluconate dehydrogenase family protein	3.76	1.61	x
AT3G62530	ARM repeat super family protein	3.71	0.94	x
AT4G31480	Coatomer	3.71	3.2	x
AT5G36230	ARM repeat super family protein	3.64	1.93	x
AT5G66420	unknown protein	3.63	1.58	
AT2G33150	PKT3	3.62	1.8	x
AT1G79530	GAPCP-1	3.58	3.1	x
AT1G24020	MLP-like protein 423	3.53	1.5	
AT1G78570	RHM1	3.49	0.98	x
AT5G11520	ASP3	3.48	2.55	x
AT3G22890	APS1	3.44	1.18	x
AT5G23860	TUB8	3.41	2.61	x

AT3G47470	LHCA4	3.37	1.45	x
AT5G13650	elongation factor family protein	3.32	2.48	x
AT4G35260	IDH1	3.32	1.02	x
AT5G09900	EMB2107	3.29	2.54	x
AT5G13780	Acyl-CoAN-acyltransferases (NAT) super family protein	3.28	3.9	
AT3G26650	GAPA	3.26	3.28	x
AT1G63000	NRS/ER	3.23	1.02	x
AT1G49430	LACS2	3.2	1.77	
AT1G12000	Phosphofructokinase family protein	3.2	5.29	x
AT1G11430	plastid developmental protein DAG	3.2	3.18	
AT5G20890	TCP-1/cpn60 chaperonin family protein	3.2	2.01	x
AT1G64520	RPN12a	3.17	2.03	x
AT3G03960	TCP-1/cpn60 chaperonin family protein	3.17	2	x
AT4G04640	ATPC1	3.15	3.7	
AT3G22960	PKP1	3.15	1.46	x
AT3G03780	ATMS2	3.14	3.92	x
AT5G45930	CHLI2	3.12	3.05	
ATCG00500	ACCD	3.09	2.39	x
AT1G32990	PRPL11	3.05	3.06	x
AT1G74960	FAB1	3.02	1.11	
AT5G20920	EIF2BETA	3.01	1.35	x
AT2G20580	RPN1A	3.01	2.67	x
AT5G43940	HOT5	2.99	2.33	x
AT4G20890	TUB9	2.95	1.24	x
AT1G61520	LHCA3	2.93	1.57	
AT2G02560	CAND1	2.91	1.75	x
AT3G07770	Hsp89.1	2.91	2.22	x
AT2G25450	2-oxoglutarate(2OG) and Fe(II)-dependent oxygenase super family protein	2.86	0.94	x
AT4G28750	PSAE-1	2.82	0.8	
AT5G03290	IDH-V	2.78	2.61	x
AT3G14420	Aldolase-type TIM barrel family protein	2.76	0.86	x
AT1G45201	TLL1	2.7	2.12	
AT1G80410	EMB2753	2.68	1.67	x
AT3G57290	EIF3E	2.66	4.11	x
AT3G18130	RACK1C_AT	2.65	1.23	x
AT2G23350	PAB4	2.64	1.19	x
AT3G18190	TCP-1/cpn60 chaperonin family protein	2.64	0.99	x
AT3G28300	AT14A	2.62	1.21	
AT3G52180	ATPTPKIS1	2.61	2.55	
AT4G16390	SVR7	2.58	3.01	x
AT1G05010	ACO4	2.57	1.35	
AT1G22410	Class-II DAHP synthetase family protein	2.56	3.34	x
AT3G11910	UBP13	2.56	2.75	x
AT1G20620	CAT3	2.55	3.17	x
AT1G50480	THFS	2.55	1.85	x
AT2G21660	ATGRP7	2.54	2.71	x
AT4G09650	ATPD	2.54	1.1	
AT1G50010	TUA2	2.53	2.98	x
AT1G08520	ALB1	2.52	1.12	x
AT1G02500	SAM1	2.5	3.27	x
AT5G04130	GYRB2	2.49	3.22	x
AT2G06850	EXGT-A1	2.48	1.71	x
AT4G04910	NSF	2.48	2.76	x
AT5G13450	ATP5	2.47	2.17	x
AT5G44090	Calcium-binding EF-hand family protein	2.47	1.55	
AT3G14990	Class-I glutamine amidotransferase-like super family protein	2.47	3.01	x
AT5G42390	Insulinase (Peptidase family M16) family protein	2.47	2.67	
AT4G15560	CLA1	2.46	3.85	
AT5G45390	CLPP4	2.46	1.03	
AT3G29320	Glycosyl transferase	2.46	1.46	
AT1G20010	TUB5	2.46	2.07	x
AT1G54080	UBP1A	2.45	2.87	
AT5G08610	P-loop containing nucleoside triphosphate hydrolases super family protein	2.44	4.43	x

AT3G13160	Tetratricopeptiderepeat(TPR)-like super family protein	2.44	1.43	x
AT5G19820	emb2734	2.37	3.69	x
AT5G18660	PCB2	2.36	2.01	
AT5G50920	CLPC	2.35	5.03	x
AT2G47730	ATGSTF8	2.32	3.51	
AT1G54270	EIF4A-2	2.31	2.66	x
ATCG00670	CLPP1	2.3	2.86	x
AT4G13930	SHM4	2.29	2.12	x
AT5G60790	ATGCN1	2.26	2.62	x
AT4G27440	PORB	2.26	3.47	x
AT3G11130	Clathrin	2.25	3.42	x
AT3G42050	vacuolar ATP synthase subunit H family protein	2.24	2.33	x
AT3G08530	Clathrin	2.22	4.71	x
AT5G42020	BIP	2.21	2.13	x
AT1G64790	ILA	2.19	3.46	x
AT4G01850	SAM-2	2.19	2.67	x
AT5G62700	TUB3	2.19	3.3	x
AT5G46290	KASI	2.15	2.98	x
AT4G39330	ATCAD9	2.14	2.45	
AT2G36880	MAT3	2.13	3.56	x
AT5G54160	ATOMT1	2.12	3.39	x
AT2G29550	TUB7	2.08	2.95	x
AT2G33040	ATP3	2.07	5.08	x
AT3G15950	NAI2	2.06	2.18	x
AT3G56150	EIF3C	2.01	2.93	x
AT1G23190	PGM3	2	3.02	x
AT1G79550	PGK	1.99	4.22	x
AT5G12250	TUB6	1.97	3.3	x
AT1G53750	RPT1A	1.96	2.64	x
AT1G05190	emb2394	1.85	2.66	x
AT3G18780	ACT2	1.81	4.26	x
AT5G19550	ASP2	1.79	4.12	x
ATCG00120	ATPA	1.74	4.44	x
AT2G38750	ANNAT4	1.72	4.06	x
AT1G09620	ATP binding aminoacyl-tRNA ligases	1.63	4.95	x

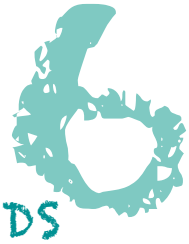
Supplemental Table S3. List of interaction partners that are considered as background as identified through pull-down on ZmAN3 in dividing or expanding leaf cells of *Zea mays*

ZmAN3		Division Zone		Expansion Zone		Previously Reported	TAP background
Protein ID	Protein Name	Difference	-LogP	Difference	-LogP		
GRMZM2G084767	Importin subunit beta-1	9.09	3.78				x
GRMZM2G005646	hypothetical protein LOC100276726	7.33	3.62	6.88	2.54	x	
GRMZM2G041765	BAG family molecular chaperone regulator 7	7.22	3.53				x
GRMZM2G091119	importin alpha-1b subunit	6.98	4.16				x
GRMZM2G138419	Osmotin-like protein	6.89	4.31			x	
GRMZM2G020146	Carboxypeptidase Y homolog A	6.83	3.2				
GRMZM2G472346	BAG family molecular chaperone regulator 7	6.12	2.44				x
GRMZM2G079668	protein DnaK	5.94	4.7				
GRMZM2G353779	BRD1	5.87	3.64	6.02	2.85	x	
GRMZM2G063864	PHD finger protein ALFIN-LIKE	5.82	3.82			x	
GRMZM2G018950	Carboxypeptidase Y homolog A	5.63	3.21				
GRMZM2G061043	pore glycoprotein p62	5.5	3.15				
GRMZM2G009845	Importin subunit alpha	5.38	2.94				x
GRMZM5G804708	ATP-dependent Clp protease proteolytic subunit	5.26	2.24				
GRMZM2G004057	60S ribosomal protein L21	5.18	3.29				
GRMZM2G088088	Importin subunit alpha	5.17	4.63				x
GRMZM2G065868	60S ribosomal protein L32	5.11	4.15				x
GRMZM2G101271	Presequence protease 1, chloroplastic/mitochondrial	4.97	3.15				
GRMZM2G042992	Topless-related protein 4	4.75	3.42				
GRMZM2G083497	40S ribosomal protein S15a	4.69	3.56				
GRMZM2G088231	importin alpha-1b subunit	4.66	2.45				x
GRMZM2G021069	minichromosome maintenance protein	4.55	3.24				
GRMZM2G020940	hydroxyproline-rich glycoprotein family protein	4.27	3.87				x
GRMZM2G083195	3-phosphate 2-O-acyltransferase	4.15	3.9				
GRMZM2G134844	ribosomal export protein NMD3	4.14	2.78				
GRMZM2G353444	Phospholipase A1-Igamma2, chloroplastic	4.05	3.57				
GRMZM2G086116	RNA helicase DHH1	3.95	2.92				
GRMZM2G127141	salt-induced MAP kinase 1	3.94	2.71				
GRMZM2G171426	40S ribosomal protein S24	3.88	3.95				x
GRMZM2G108355	PP2Ac-2	3.67	3.36				x
GRMZM2G134264	hypothetical protein LOC100275466	3.64	3.22				
GRMZM2G066101	DNA replication licensing factor mcm4	3.61	4.21				x
GRMZM2G427451	AIR12	3.59	3.4				



Mays Universalis

Roman Pleskot



A PILOT STUDY TOWARDS THE IDENTIFICATION OF DIRECT TARGETS OF ZMAN3 IN DIVIDING AND EXPANDING TISSUES

Nienke Besbrugge[1,2], Mieke Van Lijsebettens[1,2], Shubhada Kulkarni[1,2], Klaas Vandepoele[1,2], Hilde Nelissen[1,2], Jelle Van Leene[1,2] and Geert De Jaeger[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2]Centre for Plant Systems Biology, VIB, 9052 Ghent, Belgium

Author contributions

G.D.J., H.N. and J.V.L. designed the research; N.B. performed all experiments and wrote the chapter; M.V.L. provided plant material; S.K. and K.V. helped with the bioinformatics analysis; J.V.L, H.N and G.D.J. supervised and complemented the writing

ABSTRACT

In the last decade, genome-wide profiling of a binding sites of a transcription factor (TF) through ChIP-seq has become widely adopted. Besides the qualitative detection of bound regions, more and more studies are now trying to identify differences in binding profiles between two biological conditions or tissues. TAP experiments on ZmAN3 in the dividing and expanding tissues of the same growth zone showed that ZmAN3 interacts with members of the GRF TF family, and that this interaction is predominantly located in dividing cells. Based on these observed dynamics, we performed a trial experiment to identify differences in target genes of the complexes associated with AN3 in the division and expansion zones of the leaf to gain a deeper insight into the role of AN3 during leaf growth. Using the full growth zone experiment as a benchmark, qualitative differences in binding sites between both tissues were first identified. Our data suggest that the targets of ZmAN3 in the two developmental zones differ to a great extent, and that the AN3-GRF regulatory feedback loop that controls GRF (and AN3) expression is preserved between *Arabidopsis thaliana* and *Zea mays*. Furthermore, some interesting candidate target genes with a role in growth regulation were identified, although validation and further optimisation of the differential analysis will still be needed.

INTRODUCTION

In plants, many developmental processes and responses to environmental stresses are regulated at the genome level by a wide variety of DNA-binding proteins or TFs. Binding of these TFs to their target sequences generally results in changes in gene expression levels and provides an enormous flexibility as well as temporal and tissue specificity to transcriptional regulation. As such, identification of a TF's binding sites is crucial to understanding gene regulation, and helps to clarify and annotate whole functional networks. One of the most commonly applied techniques to identify binding sites of a TF in a genome-wide manner is ChIP (see Chapter 1) in combination with high-throughput NGS (ChIP-seq) (Kaufmann *et al.*, 2010; Feng *et al.*, 2017; Nolan *et al.*, 2017; Goslin *et al.*, 2017). In general, ChIP starts with the *in vivo* cross-linking of the material to preserve protein-DNA interactions. Next, the cross-linked chromatin is sheared through sonication or enzymatic digestion, and incubated with a specific antibody against the TF of interest to enrich bound DNA fragments. These fragments can then be identified in a cost and labour effective way through NGS. To circumvent the need for protein-specific antibodies, a wide variety of genetically fused tags, including the HA tag, FLAG tag and GFP tag, have been used in ChIP experiments.

Traditional ChIP-seq data analysis focusses on the qualitative detection of binding events (*i.e.*, presence or absence of a binding site) by aligning and mapping ChIP-seq reads against the genome. Peaks of high-density reads along the genome are hence called using software, such as MACS (Zhang *et al.*, 2008*b*), and represent putative binding sites. However, as NGS costs decrease, more and more

experiments are now focussing on the identification of differential binding (DB) profiles of a TF between conditions or cell types (Ross-Innes *et al.*, 2012; Pal *et al.*, 2013; Lopez-Anido *et al.*, 2015; Adriaens *et al.*, 2016), because these changes might be linked to divergent biological functions. A common strategy to assess such DB is by identifying overlapping peak regions under different conditions. Given that TF-binding signals typically provide peaks that are narrow and rather uniformly shaped, such an analyis can help identify peaks uniquely found in one condition, or peaks in common between both conditions.

In Chapter 3 we evaluated the performance of a ChIP protocol and MACS2-based (Zhang *et al.*, 2008*b*) data analysis pipeline in conjunction with the GS^{yellow} tag by using the transcriptional co-activator ZmAN3. ZmAN3 is a growth-regulating protein that was shown to associate with members of the GRF TF family and with SWI/SNF chromatin-remodelling complexes in the maize leaf, similar to its orthologue in *Arabidopsis* (Vercruyssen *et al.*, 2014; Nelissen *et al.*, 2015). More in-depth analysis of how these associations change in different developmental contexts further showed that most GRF proteins interact specifically with ZmAN3 in the division zone. To this end, high-resolution sampling of the different developmental zones (*i.e.*, division and expansion zones) to enrich for dividing or expanding cells along the maize leaf was used Here, we report on a first attempt to assess DB of the complexes associated with ZmAN3 along the genome in dividing and expanding cells at the leaf base through high-resolution sampling. Using a ZmNLS-GS^{yellow} line (see Chapter 3) as a negative control, peaks were identified via the computational workflow that was set up in the ChIP-seq experiment on the full growth zone. Although our preliminary data indicate that ZmAN3 targets the transcription of growth-regulatory TFs, such as ZmGRF11, in the division zone that are linked to the proliferative state of the tissue, future analysis is needed to confirm these findings. Here, we report on the identification of growth zone-specific target genes of ZmAN3 in maize, the encountered pitfalls and perspectives for further analysis.

RESULTS

QUALITY CHECK FOR THE IDENTIFICATION OF GROWTH ZONE-SPECIFIC TARGET GENES OF ZMAN3

We previously established a GS^{yellow}-based ChIP-seq protocol for the identification of a protein's target genes in the full growth zone of the maize leaf (see Chapter 3) using the transcriptional co-activator ZmAN3. Here we want to further fine-tune our knowledge on the working mechanism of ZmAN3 during leaf growth by comparing the target genes in the division zone with the targets in the expansion zone. To be able to identify growth zone-specific changes in the DNA binding of the ZmAN3 complex, high-resolution sampling of the different developmental zones (*i.e.*, division and expansion zones) of the leaf is needed, as well as adaptation of the experimental set-up used for the full growth zone to the

new biological input material. Therefore, a quality check of the protocol was performed using the ZmNLS-GS^{yellow} transgenic plant line.

Taking the protein input of the full growth zone ChIP-seq as a benchmark, total protein concentrations of the dividing tissue were first compared to concentrations in the expanding tissue, and were found to be approximately 2.5 times lower in the latter. Next, cross-linking of the new material through vacuum infiltration with formaldehyde was evaluated. Under-cross-linking could indicate that the target is only weakly cross-linked to the DNA, whereas over-cross-linking restricts binding of the antibody to the target protein. Cross-linking efficiency was visualised through DNA agarose gel electrophoresis of cross-linked versus reverse cross-linked samples prepared from dividing or expanding tissue (Figure 1). When the DNA is cross-linked properly, reverse cross-linking is required for efficient isolation of DNA from the nuclei. When the samples are over-cross-linked, it is impossible to recover a substantial amount of DNA by reverse cross-linking. On the other hand, when the chromatin is under-cross-linked, then the DNA can be retrieved without reverse cross-linking (Haring *et al.*, 2007). As no DNA could be detected for our cross-linked samples and reverse cross-linking recovered DNA, it was concluded that the incubation time with formaldehyde is sufficient for fixation of nuclear proteins to the DNA. Next, a time course experiment was set up on both growth zones to optimise DNA shearing through sonication. Enrichment for fragments of the desired size (200-400 bp) was obtained when the samples were sonicated for 10 min for both developmental stages (*i.e.*, 20 cycli) (Figure 1). These results show that the same ChIP experimental set-up can be used for the division and expansion zones as for the full growth zone of the maize leaf.

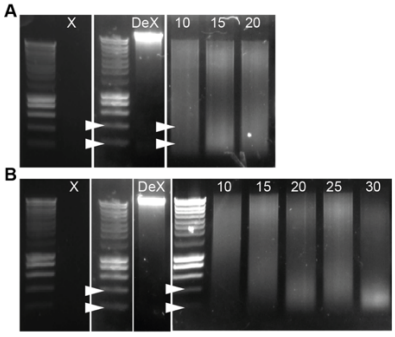


Figure 1. Optimisation of cross linking and sonication steps in the ChIP-seq protocol using DNA extracted from ZmNLS-GS^{yellow}. (A) Division zone. **(B)** Expansion zone. X: Cross-linked; DeX: Decross-linked; Numbers represent # cycli (30 s ON/OFF). Arrowheads indicate 400 and 200 bp.

CHIP-SEQ ON THE CELL DIVISION AND CELL EXPANSION ZONE OF *ZEA MAYS* LEAVES

Next, ChIPs were performed in triplicate on the dividing and expanding tissues of both ZmAN3-GS^{yellow} and ZmNLS^{yellow} leaves, using an equal protein input (5.5 mg) for all samples. All zone-specific samples represent technical replicates derived

from a single extraction (to minimise variation linked to extraction and nuclei isolation efficiency), except for the three ZmAN3 expansion zone samples. These samples were prepared in two experiments (due to too little protein for three replicates), in which the first two (A and B) originate from the same protein extract, whereas the third one (C) was derived from maize leaves that were grown and processed independently from the first two samples, making it a biological replicate (see Figure 2). Subsequent sequencing of the precipitated DNA for each sample provided a number of unique and non-redundant reads varying from 8.29E6 to 5.56E7 (Supplemental Table S1).

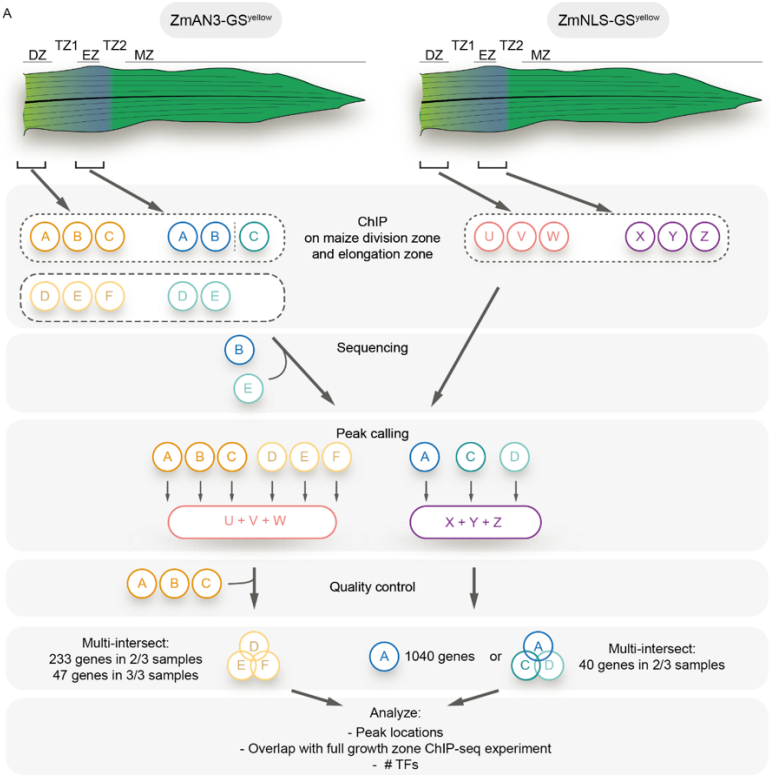


Figure 2. Schematic overview of the qualitative differential ChIP-seq analysis. Samples of the same colour were prepared from a single extract and divided as three pull downs, thus representing technical replicates. Sequencing failed for two expansion zone samples. Peak calling was carried out using a merged ZmNLS-GS^{yellow} negative control peak file for each developmental zone. The quality of individual ZmAN3-GS^{yellow} samples was assessed based on the amount of obtained peaks, peak locations and overlap of the data with a full growth zone ChIP-seq experiment. Retained samples were combined and only peaks in common between two out of three samples were used for downstream analysis concerning peak locations, overlap with the full growth zone data and the number of transcription factors. **DZ:** division zone; **EZ***:* Expansion zone; **TZ:** Transition zone; **MZ:** mature zone

For one of the three ZmAN3 expansion zone samples (sample B), sequencing failed due to the presence of sequencing adapters in the final sample. Therefore, a biological repeat experiment was carried out for the ZmAN3 samples only,

providing a second batch of three division zone samples (D, E and F; technical replicates) and two expansion zone samples (D and E, technical replicates). In this experiment, sequencing failed for one of the two expansion zone samples (E) because the amount of enriched DNA was too low (see Figure 2). For the other samples, a similar number of reads was obtained as for the first batch of samples (see Supplemental Table S1). As such, a total of six ZmAN3 division zone samples (A-B-C and D-E-F, each trio are three technical replicates) and three expansion zone replicates (A, C and D, all biological replicates) were successfully sequenced (see Figure 3).

QUALITY CONTROL OF PREPARED SAMPLES

Next, proper cross-linking of the chromatin and DNA shearing through sonication as well as binding of the fusion protein to the antibody during pull-down were assessed for a randomly selected subset of the prepared samples (Figure 3). DNA gel electrophoresis confirmed that the chromatin was cross-linked, and that reverse cross-linking yielded a substantial amount of DNA. Shearing of the DNA through sonication resulted in fragments of the desired length (*i.e.*, 200-400 bp), although the sonication efficiency clearly varied among samples (see Figure 3). For example, DNA shearing of the second batch of ZmAN3-GS^{yellow} samples showed less smear, indicating fewer longer fragments and more optimal shearing. Binding of the bait fusion protein during ChIP was evaluated by comparing the amount of fusion protein in the reverse cross-linked input sample with the amount present in the unbound fraction (FT).

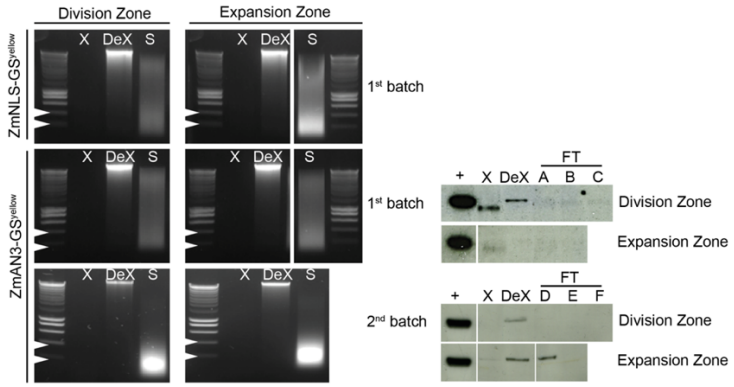


Figure 3. Validation of cross linking, sonication and binding of the bait to the antibody for both ChIP-seq batches. Left: DNA samples taken from the extracts before pull-down. **Right**: Protein samples (1/100 of the ChIP input) taken to assess proper binding of the bait (only ZmAN3-GS^{yellow} shown). X: Cross-linked; DeX: Reverse cross-linked; S: sonicated sample; FT: Flow through after binding; Arrowheads indicate 400 and 200 bp; +: positive control = 50µg extract of *Arabidopsis* cell cultures expressing 35S:AN3-GS^{yellow}

Here, efficient binding of the GS^{yellow} fusions was seen for all protein samples, except for the expansion zone sample D, in which ZmAN3-GS^{yellow} was still detected. Remarkably, some protein accumulation is seen in the cross-linked samples of the first batch, but compared to the positive control and the reverse

cross-linked samples, these are too small in size and could not reflect intact not cross-linked ZmAN3-GS^{yellow}.

QUALITY CONTROL OF SEQUENCED SAMPLES

To eliminate variation in the negative control samples, reads belonging to zone-specific ZmNLS-GS^{yellow} samples (*i.e.*, division or expanding zone) were merged prior to peak calling with MACS2 (Zhang *et al.*, 2008*b*). The merged dataset was then used as a negative control to detect peaks enriched for each ZmAN3-GS^{yellow} zone-specific sample separately (Figure 3, see Supplemental Data S1 for the raw peak data). To be able to compare the obtained data to data from the full growth zone ChIP-seq experiment, the exact same settings were used for peak detection (*i.e.*, mfold 10-30) with MACS2. Table 1 shows the number of retained peaks per ZmAN3-GS^{yellow} sample, and the corresponding number of genes as annotated in PLAZA for monocots, v3.0 (Proost *et al.*, 2015).

Table 1. Quality control of obtained peaks per sample compared to ZmNLS control samples

Sample	#Peaks	#Genes	Overlap with full growth zone ChIP-seq	
			#	%
Division Zone				
A	433	334	46	14%
B	1459	975	72	7%
C	252	189	32	17%
D	835	810	118	15%
E	342	333	47	14%
F	2520	2310	263	11%
Expansion Zone				
A	1167	1040	193	17%
C	124	97	17	14%
D	67	55	9	13%

Next, we assessed the quality of each sample and variations between the two sequenced batches by analysing the number of retained peaks, the distribution of these peaks along the genome (*i.e.*, whether they are localised in promoter, 5' UTR, 3' UTR, exonic, intronic or intergenic regions) and the overlap of the identified peaks/genes with the ChIP data derived from the full growth zone (Table 1). First of all, a high variation was observed in the number of identified peaks, ranging from 252 to 2520 peaks in the division zone and from 55 to 1040 peaks in the expansion zone. This is, however, not a consequence of the sequencing batch number or of the developmental zone from which the sample originated because a similar amount of sequencing reads was derived from all analysed samples. At the gene level and for each sample, a similar percentage (ranging from 11% to 17%) of the identified target genes overlap with the full growth zone ChIP-seq data, except for sample B of the division zone, where only a 7% overlap was observed. There is, however, a clear difference between division zone samples A-B-C (first batch), and division zone samples D-E-F (second batch): when comparing to the original fold enrichment

values of the genes identified in the full growth zone ChIP-seq, genes overlapping with the division zone samples D-E-F have higher fold enrichment values in the full growth zone ChIP than the genes that overlap with the division zone sample A-B-C, indicating a lower reliability of the latter samples.

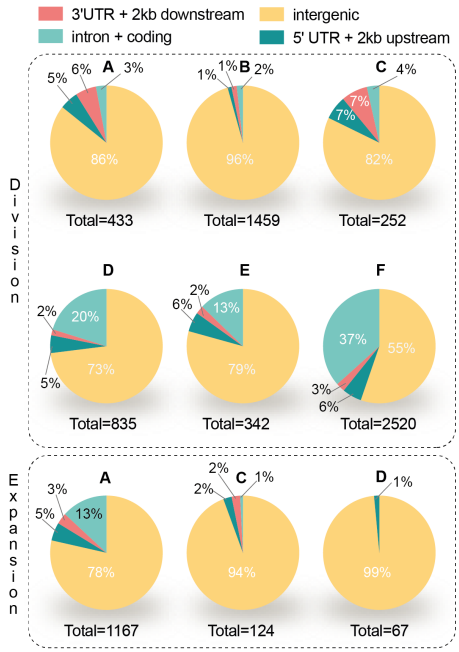


Figure 4. Distributions of peak locations in the genome for all ZmAN3-GS^{yellow} samples.

Analysis of the bound genomic regions further confirmed the lower quality of the division zone sample B, because 96% of all binding events were in intergenic regions, compared to 86% and 82% intergenic regions for the division zone samples A and C, respectively, which were from the same batch (Figure 4). Remarkably, the three division zone samples D, E and F, contained more binding events in the intron and coding region (Figure 4). A similar distribution was observed for the expansion zone sample A, whereas the expansion zone samples C and D were highly enriched for intergenic binding sites (respectively 94% and 99%) (Figure 4). On the other hand, this difference in binding distribution between the expansion zone samples might be due to the low number of identified peaks for sample C (124) and D (67). Taking these results together, we decided to continue with division zone samples D, E and F, and expansion zone samples A, C and D.

IDENTIFICATION OF BOUND GENOMIC REGIONS

To identify bound genomic regions in each growth zone, the samples were first compared with the corresponding ZmNLS negative control zone samples (as described above for the quality control). The obtained peaks are further validated based on their reproducibility among samples from the same zone and their biological relevance. As ZmAN3 was reported to influence cell proliferation during

leaf growth (Nelissen *et al.*, 2015), interesting target genes preferably have a binding site in their promoter (+5' UTR) or coding/intronic region, or encode a TF that is involved in the regulation of leaf growth. Another measure for peak validation is the overlap of the zone-specific peaks with the peaks identified in the full growth zone ChIP-seq experiment and their fold enrichment value in the latter.

Table 2. Peaks shared by at least two of the division zone samples (D, E and F) as compared to the division zone of ZmNLS

GeneID	Gene Description	Location	TLS (bp)	Rank Full Zone S1	Rank Full Zone S2	TF	Sample D FE	Sample D Q	Sample E FE	Sample E Q	Sample F FE	Sample F Q
AC209987.4	epiaristolochene 1,3-dihydroxylase	5' UTR	130	608	49				4.4	3.1	4.5	3.1
GRMZM2G017666		**2 kb**	**1467**	85	4		**5.3**	**4.1**	**5.5**	**5.1**	**5.2**	**4.5**
GRMZM2G052111	**BAM3 receptor-like kinase protein**	**intron**		37	220		**3.8**	**3.8**	**3.7**	**4.4**	**3.3**	**2.7**
GRMZM2G071089	Plasma membrane-associated cation-binding protein 1	2 kb	1088	74	142		4.4	3.2			7.4	13.3
GRMZM2G111136	SBP-domain protein1	intron		29	287	x	5.7	5.9			5.9	7.6
GRMZM2G161696	lysosomal protective protein	intron		980	544				3.6	3.4	3.9	3.6
GRMZM2G169270	SBP promoter-binding-like protein 15	2 kb	797	414	586	x	5.6	5.7			5.3	5.9
GRMZM2G176651	THICK TASSEL DWARF1	2 kb	843	218	868		5.1	3.3			5.8	7.8
GRMZM2G531804		2 kb	183	574	1132				7.5	13.2	6.9	10.9
GRMZM2G046528		2 kb	447				5.3	4.6	5.4	5.8		
GRMZM2G465091	Transcription factor PCF1 (TCP family)	2 kb	1407			x	5.9	6.5			7.0	11.1
GRMZM2G025642	NAC transcription factor NAM-B1	2 kb	1018			x	4.5	3.5			5.6	7.4
GRMZM2G066029	ZAR4	2 kb	548				4.7	2.9	5.3	4.6		
GRMZM2G067743	**GRF11**	**2 kb**	**559**			**x**	**3.6**	**3.3**	**3.2**	**2.0**	**3.4**	**3.1**
GRMZM2G069500		**2 kb**	**588**				**4.7**	**5.4**	**4.7**	**6.4**	**4.0**	**4.1**
GRMZM2G179069	Dof zinc finger protein DOF5	2 kb	491			x	3.8	1.8			4.5	3.8
GRMZM2G410150	Tetraketide alpha-pyrone reductase 2	2 kb	270				5.1	5.5			3.2	2.2
GRMZM2G441325	**ZmARF3b**	**2 kb**	**347**			**x**	**5.2**	**4.2**	**5.6**	**5.8**	**4.7**	**3.8**
GRMZM2G462760	**SAUR25 - auxin-responsive SAUR**	**2 kb**	**1299**				**17**	**32**	**3.2**	**1.7**	**3.2**	**1.7**
GRMZM5G818655		2 kb	369			x	5.1	5.5			5.0	5.9
GRMZM5G867122		2 kb	322				4.4	4.3			3.8	3.6
GRMZM2G059392	carboxy-lyase, riboside 5'-monophosphate phosphoribohydrolase	5' UTR	1654				5.4	4.7	5.4	5.5		
GRMZM2G066609		5' UTR	189				3.7	7.2			4.1	12.6
GRMZM2G108457	Prolycopene isomerase 2	5' UTR	1515						3.8	2.1	4.5	3.8
AC197387.3		coding					3.3	2.4			4.1	6.4
AC198386.3		coding					18.8	36.7			4.8	3.4
AC199541.4	Cinnamate beta-D-glucosyltransferase	coding					4.4	3.4			5.4	6.0
AC202107.3		**coding**					**5.3**	**4.6**	**6.7**	**9.5**	**4.6**	**3.9**
AC208358.3	QWRF motif-containing protein	coding							5.5	5.5	4.1	2.6
GRMZM2G031278		coding					5.0	3.9			5.0	4.5
GRMZM2G034471	cassadiene C11-alpha-hydroxylase 2	coding					3.3	1.5			4.4	3.7
GRMZM2G044128	Centromere/microtubule-binding protein CBF5	coding					4.8	4.2			4.3	3.8

GRMZM2G079619	Amorpha-4,11-diene 12-monooxygenase	coding				4.5	2.7			4.3	3.0
GRMZM2G101080		coding				3.5	1.8			4.2	4.1
GRMZM2G103197		coding				3.4	1.8			3.4	2.3
GRMZM2G115304	Protein SCAI	coding				3.9	2.7			4.0	3.5
GRMZM2G117796	CDPK-related kinase	coding						4.7	3.7	4.4	3.2
GRMZM2G124715	Trichome differentiation protein GL1	coding			x			4.4	2.8	3.6	1.7
GRMZM2G139700	NAC transcription factor NAM	coding			x	5.1	4.3			4.4	3.2
GRMZM2G148441	Amorpha-4,11-diene 12-monooxygenase	coding				2.8	1.3			4.0	3.4
GRMZM2G155216	Chlorophyll a-b binding protein Mg	coding				4.4	2.4			31.4	101
GRMZM2G156905	RLK FLORAL ORGAN NUMBER1	coding				4.5	3.9			3.3	2.1
GRMZM2G168833	Protein YLS3	coding				5.6	4.5			3.2	1.7
GRMZM2G391794	**BRI1-associated receptor kinase 1**	**coding**				**5.9**	**6.8**	**3.8**	**2.5**	**3.9**	**4.0**
GRMZM2G394450	Beta-fructofuranosidase 1 Precursor	coding				3.5	3.5			4.4	6.7
GRMZM2G400390	Laccase-12/13	coding				4.2	2.4			4.1	2.7
GRMZM2G417954	Viviparous-14	coding				3.8	1.8			3.9	2.6
GRMZM2G421463	LysM domain-containing GPI-anchored protein 1	coding				4.7	3.1			5.6	5.7
GRMZM2G428628	N-methyl-L-tryptophan oxidase	coding				3.7	1.9			4.5	3.6
GRMZM2G439195	Nicotianamine synthase, nicotianamine synthase 3	coding				4.7	3.6			4.2	3.2
GRMZM2G454334	RPP13-like protein 3	coding						4.6	3.1	7.0	9.0
GRMZM2G460371		coding						4.2	4.5	4.7	7.8
GRMZM2G479325		coding						2.8	1.4	5.4	6.4
GRMZM5G801627	CCCH domain-containing protein 35	coding			x	4.5	4.8			4.3	4.4
GRMZM5G808156		coding				4.9	4.6	5.0	6.0		
GRMZM5G880028		coding				5.0	3.7			3.4	2.1
GRMZM2G013450	Peroxisomal nicotinamide adenine dinucleotide carrier	intron						5.2	5.5	3.8	2.6
GRMZM2G017405	Tetraacyldisaccharide 4'-kinase	intron				3.9	3.0			4.2	4.7
GRMZM2G054678	Ubiquitin-like-specific protease	intron						4.3	4.4	4.6	5.5
GRMZM2G059365	box ATP-dependent RNA helicase 13	intron				4.8	3.9			4.3	3.5
GRMZM2G116292	Polyubiquitin-C	intron				4.1	4.2	3.3	2.5		
GRMZM2G131912	pectate lyase 8	intron						6.0	6.9	4.5	3.3
GRMZM2G136293	E3 ubiquitin-protein ligase ARI4	intron				4.3	4.1			3.0	2.3
GRMZM2G146446	Transmembrane protein 8B	intron						9.1	19.1	8.9	18.1
GRMZM2G386155	Serine carboxypeptidase-like	intron						4.5	4.1	5.1	5.8
GRMZM2G443495	**Glutamyl-tRNA(Gln) amidotransferase subunit A**	**intron**				**3.7**	**2.3**	**4.5**	**4.9**	**4.2**	**4.1**

Peaks retrieved in all three samples are indicated in bold.Only peaks in coding, intronic, promoter and 5' UTR regions are shown.

Location: Peak location; 2kb : 2 kb upstream region (~promoter); TLS (bp): distance to translation start site of closest gene in base pairs; Rank full zone: Ranking position of peak in overlap ZmNLS full growth zone experiment, both sample ranks (S1 and S2) are provided. Peaks were ranked according to fold enrichment; TF: Indicates with x if the gene encodes a transcription factor according to GRASSIUS and PLANTTFDBD; FE: Fold enrichment value; Q: Log(q-value) output from MACS2 (Zhang *et al.*, 2008).

To determine the conservation of peaks among samples, a multi-intersect was taken from the zone-specific samples and only peaks in common between two of the three samples were considered for further analysis (25% peak overlap).

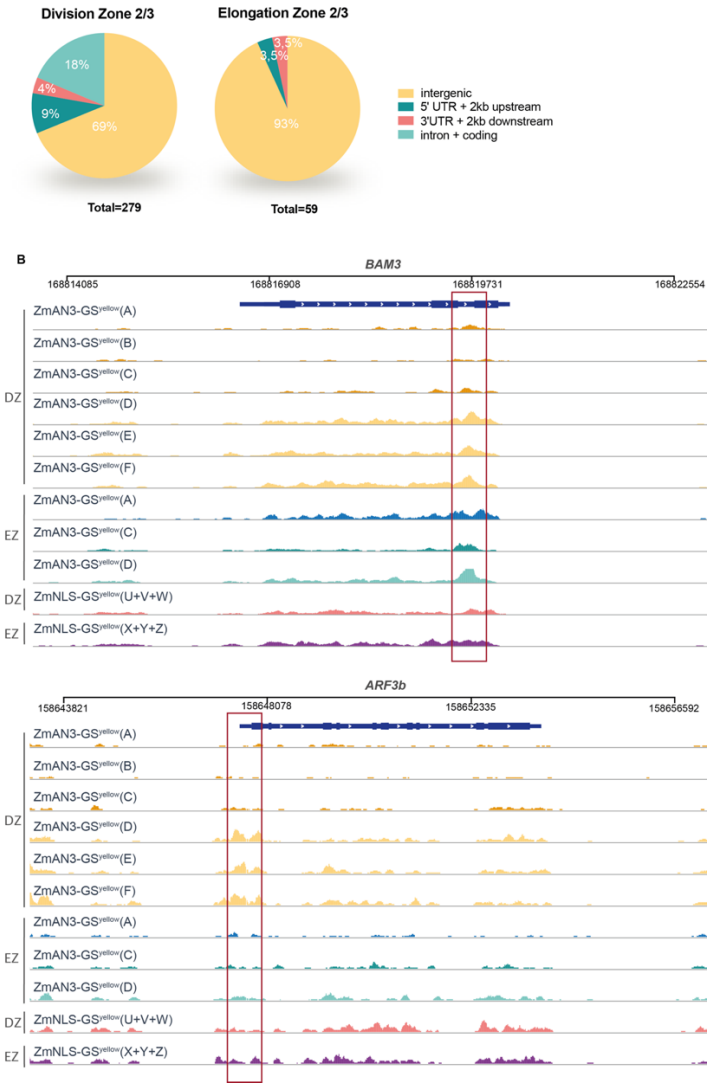


Figure 5. Validation of zone-specific data. (A) Distributions of peak locations in the genome for peaks found in at least two out of three zone-specific ZmAN3-GS^{yellow} samples. For the division zone, only sample D, E and F were considered. **(B)** Visualisation of the identified peak (red box) in the intron of BAM3 (top) and in the promoter of ARF3b (bottom) using IGV. Background levels are reproducible between samples, indicating a homogeneously sonicated chromatin extract. For the ZmNLS-GS^{yellow} control samples, reads from three replicates were merged per zone

For the division zone, this provided a total of 235 unique peaks corresponding to 233 genes of which 47 were present in all three samples. Of these 235 peaks, 9%

and 19% lie in the 2 kb upstream +5' UTR region or the coding+intron sequences, respectively, suggesting a regulatory role for ZmAN3 on these genes (Table 2, Figure 5). Interestingly, 37 peaks were also found in the full growth zone ChIP-seq, and 35 of them were amongst the 500 most enriched genes in the latter experiment. There is also clear enrichment (hypergyometric distribution; p<0.0001) for transcriptional regulators, as 33 (of which 9 were also found in the full growth zone experiment) out of the 233 genes encode a maize TF (based on Grassius TF DB and PLANTTFDBD). If we combine these selection criteria (*i.e.*, location of binding site, TF or presence in the full zone ChIP), many interesting candidate target genes remain: promoter-bound auxin-responsive SAUR25 and ZAR4, CLAVATA1 orthologue THICK TASSEL DWARF1 (TD1), TFs TCP1, ARF3b, SPL2, SPL22 and GRF11, as well as the BAM3 receptor-like kinase that was also found in the full growth zone experiment. Although the fold enrichment values of identified peaks are lower than those observed in the full growth zone ChIP-seq experiment, the specificity of the ZmAN3 peaks can still be seen by aligning the reads to the genome in the Integrative Genomics Browser (IGV), as is exemplified for the peak identified in the last intron of BAM3 (Figure 5).

Table 3. Peaks shared by at least two of the expansion zone samples (A, C and D) as compared to the expansion zone of ZmNLS

GeneID	Gene Description	Location	TLS (bp)	Rank Full Zone		TF	Sample A		Sample C		Sample D	
				S1	S2		FE	Q	FE	Q	FE	Q
GRMZM2G012923	Monosaccharide-sensing protein 3	-2 kb							64	227	21	56
GRMZM5G896363		-2 kb					4.9	43	2.5	45		
GRMZM2G176216	NAD(P)H-quinone oxidoreductase subunit 5	5' UTR	1608				4.3	32	1.9	19		
GRMZM2G463640	Cytochrome b6	5' UTR	512	2121	2120		4.4	34	1.9	15		
AC208113.3		intergenic							20	47	7.0	8.0
AC234575.1		intergenic							28	76	8.8	13.2
GRMZM2G013657	AP2-like ethylene-responsive TF	intergenic				x			27	73	6.7	7.1
GRMZM2G013798	30S ribosomal protein S2	intergenic		2125	2127		2.9	15	1.9	18		
GRMZM2G027075	**lysine-specific demethylase JMJ14**	**intergenic**		2118	2119		**3.9**	**21**	**2.2**	**213**	**1.8**	**15**
GRMZM2G027075	lysine-specific demethylase JMJ14	intergenic					3.8	25.1	1.6	6.5		
GRMZM2G027075	lysine-specific demethylase JMJ14	intergenic					16	140	3.0	41		
GRMZM2G027075	lysine-specific demethylase JMJ14	intergenic					4.0	23	2.2	25		
GRMZM2G027075	lysine-specific demethylase JMJ14	intergenic							7.4	51	6.2	45
GRMZM2G042592	Thioredoxin-like 2-1, chloroplastic	intergenic		854	1253		21	118	4.6	65		
GRMZM2G063517	Multisite-specific tRNA:(cytosine-C(5))-methyltransferase trm4b	intergenic					5.8	29	1.8	6.9		
GRMZM2G087081		intergenic							25	89	5.3	9.1
GRMZM2G096663		intergenic							24	61	4.9	2.8
GRMZM2G096952	Pleiotropic ABC efflux transporter	intergenic					5.9	29	2.4	21		

GRMZM2G096952	Pleiotropic ABC efflux transporter	intergenic					2.3	2.9	1.6	2.9		
GRMZM2G098957	protein L14	intergenic							20	59	12	27.5
GRMZM2G113158	lectin S-receptor-like serine/threonine-protein kinase	intergenic		1882	2073				23	99	8.1	23.7
GRMZM2G124441	finger protein 126-B	intergenic							19	54	4.3	6.0
GRMZM2G124593	Pyruvate kinase	intergenic							43	135	12	22
GRMZM2G171163	lysine-specific demethylase JMJ14	intergenic							3.4	30	2.7	24
GRMZM2G171163	lysine-specific demethylase JMJ14	intergenic							2.4	16	2.0	12.3
GRMZM2G174854		intergenic							2.9	4.7	2.9	6.7
GRMZM2G174854		intergenic							13	27	6.0	6.3
GRMZM2G178432	Transmembrane protein 258	intergenic		2083	1910		4.3	17.7	2.1	15.1		
GRMZM2G178432	Transmembrane protein 258	intergenic					13	131	3.0	49		
GRMZM2G317728	Pentatricopeptide repeat-containing protein At2g27800	intergenic		1403	2087		8.2	9.1	19	173	6.2	47
GRMZM2G331566	Endoglucanase 20, endoglucanase 1	intergenic							24	64	4.8	4.5
GRMZM2G332458		intergenic							28	76	10	17.6
GRMZM2G333923		intergenic							53	177	13	26
GRMZM2G356545		intergenic							79	299	25	72.7
GRMZM2G377647		intergenic							38	115	10	17.6
GRMZM2G386294		intergenic		2113	2110		6.4	71.7	2.4	40		
GRMZM2G386294		**intergenic**					**3.9**	**12**	**3.4**	**47**	**2.1**	**15**
GRMZM2G386294		intergenic					9.6	113	2.6	40		
GRMZM2G386294		intergenic					3.5	19	2.2	28		
GRMZM2G386294		intergenic					4.0	15	2.4	23		
GRMZM2G420280	DELLA protein	intergenic				x			19	52	4.9	4.3
GRMZM2G420713	Negative growth regulatory protein NGR1	intergenic					3.6	15.2	2.3	25		
GRMZM2G420713	Negative growth regulatory protein NGR1	intergenic					4.4	19.3	2.1	14		
GRMZM2G441906	Grx_A2 - glutaredoxin subgroup III, Monothiol glutaredoxin-S2	intergenic							32	100	8.0	15
GRMZM2G464272		intergenic		804	1195		13	43.1	6.4	59		
GRMZM2G476000	Putative calcium-transporting ATPase 13	intergenic							18	41	4.8	4.5
GRMZM2G523458		intergenic							7.2	19.7	8.6	36
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic		2116	2090		2.4	2.3			1.5	2.5
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic					5.5	38.7	2.8	50		
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic					3.6	18	2.4	44		
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic					9.3	130	2.4	44	2.2	11.0
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic					7.3	60	2.8	41		
GRMZM2G523971	Hypot. Mitoch. prot. mp162)	intergenic							4.9	34	2.9	17
GRMZM2G702980		intergenic							66	246	22	63.3

GRMZM5G820822	Thiol protease aleurain	intergenic					39	120	17	41
GRMZM5G851753	Putative DNA ligase 4	intergenic					27	72	8.4	12.1
GRMZM5G876160	SEC14-like protein 5	intergenic					29	81	11	20
GRMZM5G896363		intergenic					1.8	9.2	1.6	5.5
GRMZM5G897944		intergenic					41	162	15	34

Peaks retrieved in all 3 samples are indicated in bold

Location: Peak location; - 2kb : 2 kb downstream region; TLS (bp): distance to translation start site of closest gene in base pairs; Rank full zone: Ranking position of peak in overlap ZmNLS full growth zone experiment, both sample ranks (S1 and S2) are provided. Peaks were ranked according to fold enrichment; TF: Indicates with x if the gene encodes a transcription factor according to GRASSIUS and PLANTTFDBD; FE: Fold enrichment value; Q: Log(q-value) output from MACS2 (Zhang *et al.*, 2008).

For the expansion zone, 59 unique peaks corresponding to 40 genes were conserved among two out of three samples, and 93% of these peaks lie in intergenic regions (Table 3, Figure 5). Only two peaks were present in the 5' UTR and none were found in the 2 kb upstream, coding or intronic regions. In total, 12 peaks overlap with peaks identified in the ChIP-seq on the full growth zone, but only two of them were present amongst the full growth zone top 500 of most enriched genes. Although these expansion zone results might indicate that the regulatory role of ZmAN3 is restricted to the division zone, the low number of retained peaks for expansion zone-specific samples C and D could bias the identification of ZmAN3 expansion zone specific target genes.

Because the number of retained peaks and their genomic distribution in expansion zone sample A are more similar to the division zone samples D, E and F, we decided to continue with expansion zone sample A only. In total, 1167 peaks corresponding to 1040 genes were identified in this sample. About 13% of these peaks lie in the coding/intronic sequences, whereas 9% were found in the 2-kb upstream/5′ UTR regions (Table 4, Figure 4). Again, the overlap with peaks of the full growth zone ChIP-seq was determined. Remarkably, out of 1167 peaks only 16 (corresponding to 14 genes) were found in common with the full zone experiment, and all but one fell into intergenic regions. As there are no further data available to confirm any of the identified genes, it is difficult to make any conclusions on genes that have a binding site in their 2-kb upstream/5' UTR or coding/intronic regions in the expansion zone.

Table 4. Peaks identified in the expansion zone sample A as compared to the expansion zone of ZmNLS

GeneID	Gene Description	Location	TLS (bp)	Rank Full Zone		TF	Sample A	
				S1	S2		FE	Q
GRMZM2G587606	Cytochrome b6-f complex subunit 8	2 kb	1912				15.1	27.7
GRMZM2G072054	Probable succinyl-CoA ligase [ADP-forming]	2 kb	1857				5.7	4.0
GRMZM2G117851	bZIP transcription factor1	2 kb	1612				5.5	3.9
GRMZM2G541230		2 kb	1564				5.0	2.8
GRMZM2G475059	glutathione transferase31	2 kb	1545				5.2	3.2
GRMZM2G071790	Tubulin beta-8 chain,	2 kb	1419				7.4	7.3
GRMZM2G016705		2 kb	1309				7.6	7.7
GRMZM2G059392	carboxy-lyase, riboside 5'-monophosphate phosphoribohydrolase	2 kb	1290				5.9	4.4
GRMZM2G096271	Polcalcin Jun o 2, polcalcin Jun o 2	2 kb	1273				5.9	4.4
AC198982.3		2 kb	1244				5.3	3.3

GRMZM2G308689	splicing factor SYF2	2 kb	1205			6.1	4.6
GRMZM2G181607	Thioredoxin-like 1-1, chloroplastic	2 kb	1191			5.5	3.9
GRMZM2G166064		2 kb	1151			3.2	15.9
GRMZM2G054162	Probable JmjC domain-containing histone demethylation protein 2C	2 kb	1092			4.4	2.0
GRMZM2G111136	SBP-domain protein1	2 kb	1079			5.6	3.8
GRMZM2G317705	Misshapen-like kinase 1	2 kb	1026			4.1	1.5
GRMZM2G023921	DNA-binding WRKY	2 kb	1010			5.7	4.0
GRMZM5G821709		2 kb	999			8.8	9.9
GRMZM2G099382	tonoplast dicarboxylate transporter	2 kb	983			4.7	2.5
GRMZM2G425965	U-box domain-containing protein 38	2 kb	944			4.2	1.7
GRMZM2G379005	Putative GATA transcription factor 13	2 kb	903			7.8	8.4
GRMZM2G151903	toxin extrusion protein 1	2 kb	899			4.7	2.4
GRMZM2G438386	Monocopper oxidase-like protein SKU5	2 kb	835			6.9	6.0
GRMZM2G704042		2 kb	823			5.3	3.4
GRMZM2G140041		2 kb	822			5.3	3.3
GRMZM2G073535	translation factor SUI1	2 kb	764			22.8	55.1
GRMZM2G058144		2 kb	704			4.1	1.6
GRMZM2G026223	MADS-box promoter-binding-like protein 10	2 kb	698			5.7	4.0
GRMZM2G168229	Probable peptide/nitrate transporter At5g13400	2 kb	690			9.7	11.7
GRMZM2G083176	synthase delta chain	2 kb	668			5.1	2.9
GRMZM2G401040		2 kb	600			6.5	5.3
GRMZM2G318662	Probable U6 snRNA-associated Sm-like protein	2 kb	589			6.4	5.3
GRMZM2G038606	Endoglucanase 10	2 kb	571			5.8	4.1
GRMZM2G110735	Gibberellin 2-beta-dioxygenase	2 kb	553			5.5	3.6
GRMZM2G051619	CASP-like protein Os03g0817100	2 kb	549			7.5	7.2
GRMZM2G165255	remorin	2 kb	518			5.3	3.4
GRMZM2G004511		2 kb	463			6.7	6.0
GRMZM2G122357		2 kb	437			6.6	5.7
GRMZM2G070442		2 kb	382			4.7	2.4
GRMZM5G867122		2 kb	346			5.9	4.4
GRMZM2G701848		2 kb	316			5.5	3.6
GRMZM2G158359	BRI1-associated receptor kinase 1	2 kb	310			4.7	2.3
GRMZM2G448142	ATP synthase subunit beta	2 kb	295			6.2	4.9
GRMZM2G322001	PRA1 family protein, prenylated Rab receptor 2	2 kb	186			5.8	4.1
GRMZM5G809195	IAA14 - auxin-responsive Aux/IAA family member	2 kb	7			5.0	2.9
GRMZM2G463640	**Cytochrome b6**	**5' UTR**	**509**		**x**	**6.7**	**5.8**
GRMZM5G892341	30S ribosomal protein S8, chloroplastic	5' UTR	1646			6.6	5.7
GRMZM2G176216	NAD(P)H-quinone oxidoreductase subunit 5	5' UTR	1602			6.9	6.0
GRMZM5G892341	30S ribosomal protein S8, chloroplastic	5' UTR	1106			6.4	5.4
GRMZM2G358205	oxidoreductase subunit 5, chloroplastic	5' UTR	888			13.2	19.5
GRMZM2G358180	II CP43 chlorophyll apoprotein	5' UTR	463			5.5	3.8
GRMZM2G130813		5' UTR	450			4.7	4.5
GRMZM2G035584	Shikimate O-hydroxycinnamoyltransferase	5' UTR	434			6.1	4.9
GRMZM2G319786		5' UTR	361			7.5	7.3
GRMZM2G119850	BRI1-associated receptor kinase 1	5' UTR	344			11.8	18.8
GRMZM2G108865	Dof zinc finger protein DOF3	5' UTR	158			5.5	3.7
GRMZM2G098793	glycosyltransferase, EGF domain-specific O-linked N-	5' UTR	131			19.1	33.3

	acetylglucosamine transferase					
	zipper protein					
GRMZM2G116658	PROTODERMAL FACTOR 2	5' UTR	122		6.5	5.6
GRMZM2G177631	cellulose synthase8	5' UTR	116		5.5	3.7
GRMZM2G374986	Putative Myb family transcription factor	5' UTR	105		6.9	8.6
GRMZM2G327059	BEL1-related homeotic protein 14, BEL1-like homeodomain protein	coding			5.4	3.4
GRMZM2G374986	Putative Myb family transcription factor At1g14600	coding			4.7	2.4
GRMZM2G463904	BRI1-associated receptor kinase 1	coding			5.1	2.9
GRMZM2G038152	Tip elongation aberrant protein 1	coding			7.5	7.3
GRMZM2G169825	POZ domain-containing protein At5g66560	coding			4.8	2.5
GRMZM2G093809	BRI1-associated receptor kinase 1	coding			4.9	2.7
GRMZM2G091578	LOC100285096, Xylem serine proteinase 1	coding			5.6	3.8
GRMZM2G144172	Dof zinc finger protein DOF5	coding			5.5	3.6
GRMZM2G396553	Biosynthetic arginine decarboxylase	coding			5.6	4.0
GRMZM2G119287		coding			5.0	2.9
GRMZM2G039757	Protein TolB	coding			5.5	3.7
GRMZM2G302293	Transcription factor	coding			10.5	13.6
GRMZM2G074032	finger protein JACKDAW	coding			5.1	2.9
GRMZM2G476538	IAA-amino acid hydrolase ILR1-like	coding			6.0	4.5
GRMZM2G339540	Putative inactive receptor-like protein kinase	coding			4.8	2.5
GRMZM2G056252	Omega-6 fatty acid desaturas	coding			6.5	5.4
GRMZM2G160466		coding			6.8	6.5
GRMZM2G399641	E3 ubiquitin protein ligase	coding			5.7	4.0
GRMZM2G098793	glycosyltransferase	coding			10.1	13.2
GRMZM2G057483	Anthranilate N-benzoyltransferase protein	coding			5.9	4.4
GRMZM2G121312	THICK TASSEL DWARF1	coding			5.5	3.9
GRMZM5G858456	fucosyltransferase 7, Probable fucosyltransferase	coding			9.4	12.1
AC207342.3		coding			5.7	4.0
GRMZM2G010372	WALLS ARE THIN 1	coding			5.3	3.4
GRMZM2G097404	QWRF motif-containing protein	coding			4.3	32.0
GRMZM2G039725	S-ribonuclease binding protein SBP1	coding			3.8	22.7
GRMZM2G446213		coding			2.6	9.6
GRMZM2G476699	mannose beta-1,4-N-acetylglucosaminyltransferase 2	coding			8.2	8.6
GRMZM2G041141	U-box domain-containing protein 25	coding			4.4	1.9
GRMZM2G059693	Feruloyl esterase A	coding			7.7	7.8
GRMZM2G331638	dopamine beta-monooxygenase, chelate reductase 1	coding			4.9	2.7
GRMZM2G044824	Trichome differentiation protein GL1	coding			4.7	2.3
GRMZM2G004445		coding			5.9	4.3
GRMZM2G074317	Putative cyclic nucleotide-gated ion channel	coding			5.3	3.3
GRMZM2G067915	Probable carboxylesterase 18	coding			5.3	3.4
GRMZM2G381576	cell Division Protein AAA ATPase family	coding			5.5	3.5
GRMZM2G339909		coding			9.2	15.6
GRMZM2G392116		coding			5.5	3.6

GRMZM2G016637	Trihelix transcription factor GT-3b	coding			6.4	5.1
GRMZM2G060720		coding			5.5	3.6
GRMZM2G080524	diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2	coding			10.2	16.1
GRMZM2G162508	fatty acid elongase, 3-ketoacyl-CoA synthase	coding			6.6	5.9
GRMZM2G076307	Probable glucuronosyltransferase	coding			5.2	3.1
GRMZM2G181219		coding			5.5	3.7
GRMZM2G126196	Expansin-A4	coding			6.2	4.9
GRMZM5G831102	gibberellin receptor GID1L2	coding			4.9	2.6
GRMZM2G048651	THICK TASSEL DWARF1	coding			6.7	5.8
GRMZM5G809195	IAA14 - auxin-responsive Aux/IAA family member	coding			8.3	8.9
GRMZM2G013530	Agmatine coumaroyltransferase-2	coding			6.5	5.6
GRMZM2G006197		coding			5.3	3.4
GRMZM2G072808	Laccase-12/13, L-ascorbate oxidase	coding			5.3	3.4
GRMZM2G108714	debenzoyl-2'-deoxytaxol N-benzoyltransferase,	coding			6.1	4.9
GRMZM2G044493		coding			7.1	6.8
GRMZM2G380088	Cytidine deaminase, CDA1	coding			7.1	6.8
GRMZM2G116063	Putative wuschel homeobox protein	coding			5.2	3.2
GRMZM2G135362	PH and SEC7 domain containing protein secG	coding			4.4	2.3
GRMZM2G132251	Transcription factor	coding			11.9	16.4
GRMZM2G319947		coding			6.5	5.6
GRMZM2G087254	adenylylsulfate reductase 1	coding			5.2	3.1
GRMZM2G015709	Anthocyanidin 3-O-glucoside 2''-O-glucosyltransferase	coding			6.1	4.9
GRMZM5G837822	hevamine-A, Acidic endochitinase	coding			5.5	3.6
GRMZM2G125241	endo-1,3-beta-glucosidase-like protein	coding			5.9	4.4
GRMZM2G165914	hydroxybenzoic acid efflux pump subunit AaeB	coding			6.4	5.1
GRMZM2G064903	DNA-binding WRKY	coding			4.8	2.5
GRMZM2G386209	Auxin-induced in root cultures protein 12	coding			4.7	2.4
GRMZM2G704330	LOB domain-containing protein	coding			5.7	4.0
GRMZM2G023285		coding			5.8	4.4
GRMZM2G027068	Putative transcription factor bHLH107	coding			5.0	2.9
GRMZM2G111136	SBP-domain protein1	coding			5.2	3.1
GRMZM2G146614	CONSTANS-LIKE	coding			6.5	5.6
GRMZM2G006197		coding			6.1	4.6
GRMZM2G010987	a, Anthocyanidin 5,3-O-glucosyltransferase	coding			7.1	6.8
GRMZM2G069335		coding			5.5	3.5
GRMZM2G416622		coding			5.1	2.9
GRMZM2G067865	BOI-related E3 ubiquitin-protein ligase 3	coding			10.8	25.8
GRMZM2G096655		coding			5.8	4.4
GRMZM2G089231		coding			5.3	3.3
GRMZM2G386154	protein ClpB	coding			6.3	5.0
GRMZM2G478396	Putative wuschel homeobox protein	coding			5.2	3.0
GRMZM5G862956	harpin-induced protein	coding			8.2	8.7
GRMZM5G827174		coding			5.5	3.5
GRMZM2G105834	Homeobox-leucine zipper protein	coding			8.9	10.3
GRMZM2G323387		coding			5.0	2.7
GRMZM2G177046	Ocs element-binding factor 1	coding			5.0	2.8
GRMZM5G851485	Transcription factor	coding			5.5	3.7
GRMZM2G007306		coding			6.1	4.7
GRMZM2G174784	AP2-like ethylene-responsive transcription factor	coding			4.4	34.2

GRMZM2G413647	hydroxyisobutyrate dehydrogenase-like	coding					5.5	3.9
GRMZM2G093195	Cytochrome P450 88A1(Dwarf3 protein)	coding					4.4	2.0
GRMZM2G000739	Siroheme synthase	coding					11.3	15.0
GRMZM2G123909	GATA transcription factor 20	coding					5.5	3.7
GRMZM2G059706	Dirigent protein, disease resistance response protein 206	coding					5.8	4.1
GRMZM2G393349	H2 finger protein, protein binding protein	coding					4.8	2.5
GRMZM2G021909	Probable clathrin assembly protein	coding					5.5	3.9
GRMZM2G392106		coding					7.5	7.3
GRMZM2G041761	plant-specific domain TIGR01568 family protein	coding					5.0	2.9
GRMZM2G096904		coding					6.4	5.4
GRMZM2G050912		coding					8.4	9.1
AC191071.3	Putative flavin-containing monooxygenase 2	coding					7.0	6.4
GRMZM5G852801	BES1/BZR1 homolog protein	coding					16.0	34.5
GRMZM2G036340	Gibberellin 3-beta-dioxygenase 2-2	coding					4.4	1.9
GRMZM2G171464	Transcription factor	coding					5.7	4.0
GRMZM5G898755	Non-specific lipid-transfer protein	coding					7.1	6.8
GRMZM5G846152		coding					3.9	1.4
GRMZM2G475683	induced protein	coding					6.4	5.1
GRMZM2G132303		coding					7.1	6.8
GRMZM2G131340		coding					10.8	16.7
GRMZM2G374313	Carbonyl reductase [NADPH] 1	coding					5.3	3.3
GRMZM2G022629		coding					5.0	2.9
GRMZM2G001514	Fasciclin-like arabinogalactan protein	coding					4.3	26.1
GRMZM2G070797	finger Ran-binding domain-containing protein 2	coding					5.4	24.0
GRMZM2G339736	Aspartic proteinase PCS1	coding					5.8	4.1
GRMZM2G162065	Exocyst complex component EXO70A1	coding					5.2	3.1
GRMZM2G082372	wiscott-Aldrich syndrome protein	coding					5.4	3.4
GRMZM2G433162	Probable amino acid permease 7	coding					5.7	4.0
GRMZM2G029879	Cyclin-U4	coding					5.3	3.4
GRMZM2G107711		coding					5.1	2.9
GRMZM2G087059	Ethylene-responsive transcription factor	coding					4.1	1.6
GRMZM2G144224	Abscisic acid receptor	coding					8.2	15.9
GRMZM2G153359	Gibberellin 2-beta-dioxygenase 8	coding					6.5	5.4
GRMZM2G334084		coding					5.1	2.9
GRMZM2G308595	Nudix hydrolase 21,	coding					7.4	7.5
GRMZM2G144155	Probable methyltransferase, ankyrin protein kinase-like	coding					6.2	5.9
GRMZM2G030038		coding					5.4	3.4
GRMZM2G396553	Biosynthetic arginine decarboxylase	coding					7.2	6.7
GRMZM2G055243	Homeobox protein knotted-1-like	coding					6.3	5.1
GRMZM2G046911		coding					8.5	9.3
GRMZM2G025642	NAC transcription factor NAM-B1	coding					6.3	5.4
GRMZM2G079683		coding					7.4	7.2
GRMZM2G126199	Short-chain dehydrogenase TIC 32, chloroplastic, retinol dehydrogenase 12	intron					6.1	4.9
GRMZM2G041878		intron		1420	1665		4.7	2.4
GRMZM2G028114	Probable cytosol aminopeptidase	intron					7.5	7.3
GRMZM2G035999	light-mediated development protein DET1	intron					7.8	8.8

GRMZM2G482657		intron				4.0	21.3
GRMZM2G023625	Vegetative incompatibility protein HET-E-1	intron				6.7	5.9
GRMZM2G146047	Probable ATP-dependent RNA helicase spindle-E	intron				8.2	8.6
AC198418.3	box ATP-dependent RNA helicase 3, chloroplastic	intron				8.0	8.2
GRMZM2G084014	TRANSPARENT TESTA 1	intron				7.5	7.3
GRMZM2G076398	Probable alpha-galactosidase D	intron				5.1	2.9
GRMZM2G000264	plasma membrane-type	intron				7.7	8.0
GRMZM2G115757	FK506-binding protein 2	intron				5.5	3.6
GRMZM2G374986	Putative Myb family transcription factor A	intron				6.4	5.2
GRMZM2G126260	Auxin efflux carrier component	intron				5.5	3.6
GRMZM2G062885	CHLOROPLAST IMPORT APPARATUS 2	intron				5.5	3.7
GRMZM2G139965		intron				8.9	10.1
GRMZM2G086573	AP2-like ethylene-responsive transcription factor	intron				6.2	4.8
GRMZM2G122277	Cellulose synthase A catalytic subunit	intron				5.7	4.4
GRMZM2G052111	BAM3 receptor-like kinase	intron				6.1	4.6
GRMZM2G171871	lipid-associated protein 12, chloroplastic	intron				5.2	34.6
GRMZM2G043854	Transcription factor bHLH30	intron				5.7	4.3
GRMZM2G089836	fructosyltransferase/2,1-fructan	intron				6.9	6.0
GRMZM2G460012	Vesicle-associated protein 1-1	intron				5.3	3.4
GRMZM2G083243	Tubulin alpha-6 chain, Tubulin alpha-5 chain	intron				6.4	5.3
GRMZM2G346455	Probable alpha-galactosidase D	intron				5.4	3.6

Peaks retrieved in all 3 samples are indicated in bold.
Location: Peak location; 2kb : 2 kb upstream region; TLS (bp): distance to translation start site of closest gene in base pairs; Rank full zone: Ranking position of peak in overlap ZmNLS full growth zone experiment, both sample ranks (S1 and S2) are provided. Peaks were ranked according to fold enrichment; TF: Indicates with x if the gene encodes a transcription factor according to GRASSIUS and PLANTTFDBD; FE: Fold enrichment value; Q: Log(q-value) output from MACS2 (Zhang *et al.*, 2008).

DISCUSSION AND CONCLUSION

EVALUATION OF THE CHIP-SEQ PROCEDURE

The identification of target gens of a TF through ChIP-seq has become one of the standard techniques in the functional genomics field. As the reduction in NGS costs makes it feasible to analyse multiple samples at the same time, more and more comparative ChIP-seq experiments are being carried out (Ross-Innes *et al.*, 2012; Pal *et al.*, 2013; Lopez-Anido *et al.*, 2015; Adriaens *et al.*, 2016). However, assessing the differential binding of a TF to its target genes is still very challenging and requires the integration of statistical methods to minimise false positives (Robinson *et al.*, 2009; Lun and Smyth, 2014,*a*,*b*; Allhoff *et al.*, 2016). Here, we evaluated a comparative ChIP-seq analysis on the division and expansion zones of a *Zea mays* leaf during steady-state growth, using the ZmAN3 transcriptional co-activator as a bait. Crucial to every ChIP procedure is the optimisation of cross-linking and DNA shearing through sonication efficiency before carrying out the final experiment (Haring *et al.*, 2007). Our data suggest that the implemented cross-linking time was sufficient for chromatin fixation, as a substantial amount of DNA

could be recovered only when the samples were reverse cross-linked. This is in line with publicly available protocols for ChIP in maize (Lakhal *et al*), although it has been proposed that cross-linking of co-activators, such as ZmAN3, requires longer incubation times with formaldehyde or even two cross-linking steps (Tian *et al.*, 2011).

QUALITY CONTROL SHOWS HIGH VARIABILITY BETWEEN SAMPLES

One of the main issues with ChIP experiments is the low reproducibility between samples (Landt *et al.*, 2012; Devailly *et al.*, 2015), which is influenced by many factors including cross-linking and DNA fragmentation efficiency, bait expression level, and affinity and specificity of the antibody used during the IP. Because of this low reproducibility, at least two biological replicates are generally required to confirm identified target genes (Landt *et al.*, 2012). Differences in sonication efficiency between biological repeats is one of the factors that influence reproducibility. However, even between technical replicates (where the sonication should be uniform) peak reproducibility was rather low. When the raw sequencing reads are visualised against the genome, a semi-constant pattern of background/noise reads is observed among all samples. The pattern shows regions with high background levels and regions completely depleted of sequencing reads, possibly reflecting the condensation status of the DNA representing either hetero- or euchromatin (see Figure S1). The conservation of the background pattern indicates that both the extracted input chromatin and the non-specific binding of chromatin during the ChIP are not responsible for the observed variability. Moreover, similar amounts of reads were obtained for the division zone samples D-E-F, indicating that the variance in number of peaks is not due to variations in sequencing efficiency. As such, the differences between the three samples are most probably due to variations in pull-down efficiency. Possible explanations for this variability are differences in epitope accessibility within fixed chromatin or distinct binding characteristics of the bait to the DNA. As ZmAN3 does not have a DBD of its own, the DNA-binding affinity of the TFs that interact with ZmAN3 can cause a lot of variation.

Besides biological repeats, the reproducibility of identified peaks can also be assessed using technical replicates or by analysing the overlap with known target genes (Landt *et al.*, 2012). As such, peaks in common between at least two of the three replicates were here determined for the division zone samples D-E-F, which resulted in 233 genes of which 47 were present in all three repeats. Analysis of the overlap between these 233 genes and targets identified in the full growth zone ChIP experiment provided 37 genes in common. Interestingly, 95% of the overlapping genes were found in the top 500 most enriched genes of the full growth zone experiment, showing a high reliability of this specific subset. Furthermore, we observed that the overall fold enrichment values of the full growth zone ChIP were significantly higher than those for the division zone samples D-E-F, indicating a higher efficient enrichment of ZmAN3-bound chromatin fragments in the full growth zone. Possibly, the lower fold enrichment values in the new experiments are

caused by the shorter cross-linking time, or because the expression levels of ZmAN3 have decreased during the generation of new seed stocks, although there is no real evidence to support this. Comparison to the full growth zone data also further confirms the lower quality of the samples A, B and C, because the targets in common with the full growth zone experiment had lower fold enrichment values in the latter. Finally, as TFs are thought to bind mostly to promoter and coding/intronic regions, peak location is also a good parameter to identify true target genes and assess sample quality. Interestingly, it was recently shown that intronic regions can have a highly significant effect on the level of gene expression, and that the position of the intron can have an effect on determining the position of the TLS (Gallegos and Rose, 2017). Unfortunately, the maize genome in PLAZA is still not completely annotated. Hence, peaks located in intergenic regions at a great distance from the nearest gene (> 10-20 kb) could represent *bona fide* binding events linked to nearby not yet annotated protein-coding genes, long non-coding RNAs, or miRNAs. In addition, binding in intergenic regions might reflect binding of the AN3 complex to enhancer regions, which are often located faraway from the genes that are regulated.

For the expansion zone samples, sequencing failed for two out of five analysed samples. To be able to validate the identified peaks in the remaining samples (biological replicates), the same strategy was used as for the division zone samples. Strikingly, only a poor fraction of peaks was shared by at least two of the three samples, and most of them had an intergenic location. The overlap with the full zone ChIP-seq data was also low, and from the overlapping peaks, only two were found amongst the top 500 most enriched ones in the full growth zone experiment. No peaks were located in the promoter regions, further confirming the poor quality of the expansion zone samples. One possible explanation for the low overlap between the expansion samples could be the use of biological replicates instead of technical replicates, which are inherently linked to an even higher variation compared to technical replicates, because variation increases due to differences in biological material, cross-linking, chromatin extraction and DNA fragmentation. The fact that sequencing failed for two samples that represent technical replicates of a sample that was successfully sequenced, suggests that variance in pull-down efficiency is the main driver for variation. Another possible explanation for the low success rate of AN3 ChIP in the expansion zone, is the lower concentration of ZmAN3 in the expansion zone than in the division zone, which might provide insufficient amounts of enriched DNA for proper target analysis. However, it might simply be that this poor quality comes from the fact that ZmAN3 is biologically inactive in the expansion zone, as the *an3* mutant in *Arabidopsis* was found to be mostly affected in leaf cell number and not in cell size, indicating a more prominent role in the division zone. Furthermore, ZmAN3 is thought to associate with GRFs to bind the DNA, and all ZmGRFs, except ZmGRF10, are expressed only in the dividing tissue. Although our previously reported TAP data on ZmAN3-GS^{yellow} in the different growth tissues of the maize leaf showed that ZmGRF10 associates with ZmAN3 in the expansion zone, reverse TAP on ZmGRF10 failed to confirm this interaction. Rather, ZmGRF10 was found to interact with ZmGIF2 and ZmGIF3 (data

not shown) in TAP, whereas Y2H analysis only showed direct interactions with ZmGIF1 and ZmGIF2 (Wu *et al.*, 2014). As such, ZmAN3 might not bind the DNA specifically in the expansion zone, and thus the negative results of the expansion zone might be biologically relevant.

ZMAN3 BINDS IN THE DIVISION ZONE TO REGULATORY SEQUENCES OF GENES INVOLVED IN LEAF DEVELOPMENT

From the division zone data, we were able to select a handful of candidate target genes with high confidence. Although there was no time to validate any of these candidate targets, the possible role of some interesting genes is explained only through what is found in the literature.

The first target identified in dividing cells is *ZmGRF11*, which is bound by ZmAN3 in the 2 kb upstream of the translation initiation site. Interestingly, differential TAP on maize leaves using ZmAN3-GS^{yellow} as bait showed how both proteins interact in dividing cells specifically (see Chapter 3). Moreover, ZmGRF11 expression was shown to be high in the division zone, and mRNA accumulation rapidly declines as cells go into the transition and expansion phase thanks to post-transcriptional regulation of miR396a (Zhang *et al.*, 2008*a*; Candaele *et al.*, 2014; Nelissen *et al.*, 2015). As such, because a mechanism of positive feedback regulation of GRF expression through AN3/GRF was previously reported in *Arabidopsis* (Hoe Kim and Tsukaya, 2015), it seems that the AN3/GRF growth-controlling network is very robust, complex and multi-layered, involving dynamic protein complexes, positive transcriptional feedback loops and negative feedback through miRNA-mediated post-transcriptional regulation in both monocots and dicots. Remarkably, ChIP on the full growth zone did not pick up *ZmGRF11* as a target gene, which suggests that high-resolution sampling of the division zone provides a more in-depth view of the regulatory function of ZmAN3. However, it could also indicate that *ZmGRF11* is a false positive result.

Other factors that were bound in the promoter region and that were not identified in the full growth zone ChIP-seq are the auxin-inducible SAUR25, ZAR4, and ARF3b proteins, which have their peak summits at respectively 1299, 548 and 347 bp upstream of the translation start site (TLS). SAUR25 is a Small Auxin-up RNA (SAUR) that belongs to the *SAUR* family of 79 genes, all of which have an early response to auxin (Chen *et al.*, 2014). In *Arabidopsis* they are mostly expressed in elongating tissue (Spartz *et al.*, 2012), and SAUR25 is predicted to interact with expansins in *Zea mays* according to the string-db database (http://string-db.org/). As cells are still actively dividing in the division zone, binding of ZmAN3 might reflect negative regulation of SAUR25 transcription to prevent the cells from elongating too early. However, expression levels of *SAUR25* were reported to be high at the base of the leaf, and to decrease gradually along the blade of the leaf. This suggests that binding of ZmAN3, which shows a similar expression pattern in the leaf, has a positive effect on SAUR25 expression and indicates that SAUR25 might play a specific role in cell proliferation instead of in cell expansion. It might also suggest the presence of a feedback mechanism to control SAUR25expression.

ZAR4 on the other hand, is a member of the *Zea mays* ARGOS protein family of nine, that are orthologues of the ARGOS (Auxin Regulated Gene involved in Organ Size) protein in *Arabidopsis* (Shi *et al.* 2016). ZAR proteins were reported to improve yield, organ growth and stress tolerance. Although not much is known about the function of ZAR4, *ZAR1* was recently reported to positively control organ size through changes in cell number linked to a faster growth rate. Transgenic plants overexpressing *ZAR1* were higher, had larger organs, and were more tolerant against drought stress (Guo *et al.*, 2014). However, it is still not clear if and how the link between ARGOS and auxin that is present in *Arabidopsis*, is preserved in maize. The ZmARF3b protein, an orthologue of *Arabidopsis* AUXIN RESPONSE FACTOR3/ETTIN, is known to be involved in spatiotemporal patterning during plant development. It was previously shown that ARF3 accumulation is limited to the abaxial side of the leaf by the adaxial expression of *TAS3*, which produces a tasiR-ARF that targets the mRNA of ARF3 and ARF4. In maize, mRNA of *ARF3b* has a single target site for tasiR-ARF, which is produced in the leaf through RAGGED SEEDLING2 (RGD2) and LEAFBLADELESS1 (LBL1) (orthologues of *Arabidopsis* ARGONAUTE7 and SUPPRESSOR OF GENE SILENCING3). Plants carrying a mutation in *RGD2* or *LBL1* were found to produce severely defective leaves (Dotto *et al.*, 2014). Moreover, RGD2 has been reported to regulate mediolateral expansion without influencing abaxial/adaxial leaf polarity in maize leaves (Douglas *et al.*, 2010). This suggests that ZmARF3b might also be involved in mediolateral expansion rather than dorso-ventral polarity. In *Arabidopsis*, *ARF3* expression is additionally regulated at the transcriptional level by the AS1-AS2 complex, which represses the promoter of *ARF3* through binding of AS1 (see Figure 6).

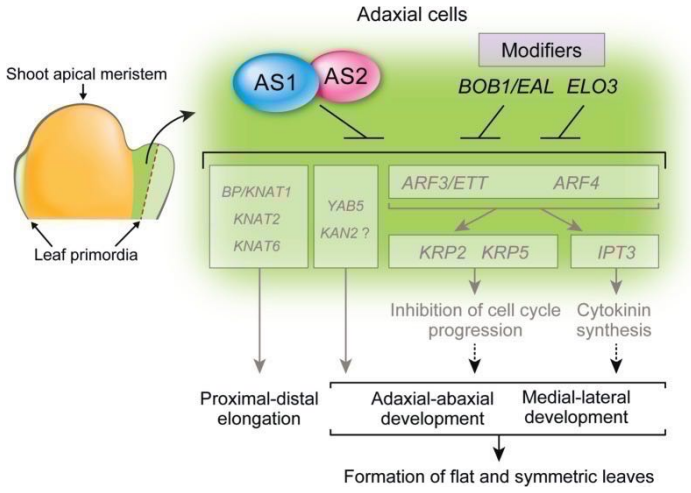


Figure 6. Schematic representation of ARF3 regulation and function in *Arabidopsis*. Figure taken from (Takahashi *et al.*, 2013)

Interestingly, combining mutations in *as1* or *as2* with *an3* results in more severe leaf polarity defects (Horiguchi *et al.*, 2010). Similarly, overexpression of miR396 in *as1* or *as2* mutant backgrounds also enhances the leaf phenotype, indicating that

the effect of the added *an3* mutation happens through interaction with GRFs (Wang *et al.*, 2010). Interestingly, it has been suggested that ARF3 positively regulates the expression of the cyclin-dependent kinase inhibitors KRP2 and KRP6, which inhibit progression through the mitotic cell cycle, and correct cell proliferation was shown to be necessary to establish correct patterning during leaf growth (Takahashi *et al.*, 2013). According to its function in leaf growth, *ZmARF3b* is mostly expressed at the base of the leaf and its expression decreases slowly over the length. As maize contains four other ZmARF3 proteins (Dotto *et al.*, 2014), a deeper investigation is needed to establish the role of ZmARF3b and ZmAN3 in leaf development.

Two other interesting candidate targets are members of the SQUAMOSA promoter-binding protein-like family of 31 TFs, namely SPL2 and SPL22, which were also found in the full growth zone ChIP-seq experiment. Conserved among plants, SPL TFs play an important role in plant growth and development and are expressed in various tissues in maize (Mao *et al.*, 2016). The specific role of each individual factor is, however, still unknown. Here we found that SPL2 and SPL22 are bound respectively in the intron and in the coding region of the gene. Interestingly, it has recently been proposed that binding to intronic regions in *Arabidopsis* might have a greater influence on the determination of the transcription start site and the expression level of the mRNA than promoter regions (Gallegos and Rose, 2017). As such, the role of ZmAN3 binding is an interesting feature to look into, especially because ZmSPL22 was found to be present in the top20 most enriched genes in the full growth zone ChIP-seq experiment.

Another interesting candidate target gene is TCP1 (TEOSINTE BRANCHED 1, CYCLOIDEA, PCF1 (TCP) 1), a TF of the TCP family, which are known regulators of leaf growth in *Arabidopsis* (Li, 2015). TCP1 belongs to the class II TCP TFs, and its expression has been reported in maize to correlate with leaf size (Baute *et al.*, 2016). Interestingly, TCP1 shows the same expression pattern in the leaf as ZmAN3, suggesting that ZmAN3 positively regulates TCP1.

Finally, our analysis reconfirmed the BARELY ANY MERISTEM3 (BAM3) and THICK TASSEL DWARF1 (TD1) genes that were also found in the full growth zone ChIP experiment, and are bound in the intronic and promoter regions, respectively. TD1 is the maize orthologue of CLAVATA1 (CLV1) in *Arabidopsis*, and is a known regulator of meristem identity in the SAM. Remarkably, *td1* mutant maize plants were found to produce fewer leaves than WT plants (Bommert, 2005), which indicates that its SAM restricted function in *Arabidopsis* is extended to leaf growth in maize. BAM3 on the other hand is a CLAVATA1-like protein, but its function in maize is presently unknown. In *Arabidopsis,* BAM3 was shown to be partly functionally redundant to CLV1, but has a less restricted expression profile in the SAM. Moreover, *bam1 bam2 bam3* triple mutants showed smaller leaf phenotypes with a reduced number of higher order nerves, suggesting a regulatory role of BAM3 in the control of the final leaf shape and in the development of the leaf vasculature (DeYoung *et al.*, 2005). Interestingly, both BAM3 and TD1 show similar expression levels at the leaf base, but, whereas *TD1* expression, such as *ZmAN3,* decreases fast, *BAM3* expression decreases slowly along the leaf length and remains present also

at the leaf tip (Sekhon *et al.*, 2013). As such, it is unclear whether ZmAN3 regulates both genes during leaf growth in a positive or a negative manner.

In summary, it is clear that many interesting genes with a link to leaf development are present in the final filtered dataset, further sustaining the reliability of the identified targets. Moreover, this shows the complexity of the underlying gene regulator networks involved in maize leaf development. It will be interesting to confirm any of these selected candidate target genes in the future to get a deeper insight into the regulatory role of ZmAN3 binding to these genes. Using for example an inducible ZmAN3 line in combination with RNA-seq might help understand whether ZmAN3 represses or activates the selected target genes. Another idea is to directly perform ChIP on the ZmGRFs that associate with ZmAN3, because these might have a smaller amount of potential binding sites. However, the instability of the ZmGRFs due to miR396-mediated degradation makes it hard to perform experiments on the native ZmGRFs. By zooming in on the division zone only, where miR396 levels are still low, using a miR396-resistant ZmGRF could help to identify target genes without creating a perturbed situation in which ZmGRFs are still present in the developmental context where they normally do not occur.

In conclusion, we reported here on the successful application of our GS^{yellow} based ChIP-seq protocol to discover growth zone-specific target genes of the ZmAN3 transcriptional co-activator. The obtained data suggest that ZmAN3 binds to the regulatory sequences of proteins that play a role in the maintenance of the proliferation activity and establishment of leaf polarity in the division zone of the leaf, but does not play a vital transcriptional regulatory role in the expansion zone.

MATERIALS AND METHODS

PLANT MATERIAL

Growth and harvesting of the transgenic lines used for ChIP-seq was carried out as described in Chapter 3.

Western blot Analysis

Protein samples were taken at several stages of the ChIP to validate binding of the bait to the antibody during ChIP. Samples were loaded on 12% Mini-PROTEAN®TGX™ Precast Protein Gels (Biorad©). Blotting and development of the blots was carried out as described (Van Leene *et al.*, 2014).

CHIP-SEQ

For each *Zea mays* division zone sample, 0.8 g of tissue was fixed on ice in buffer A (10 mM Tris/HCL, pH 8, 400 mM sucrose, 1 mM EDTA, 1 mM PMSF, 1% formaldehyde; maximum 1 g per 10 mL) for 10 min under vacuum conditions. For each expansion zone sample, 2 g of biological tissues was used (divided into 2 to make sure 1 g was submerged in 10 mL buffer A). Next, glycine was added to a final concentration of 0.1 M to stop the reaction, and incubation was continued under

vacuum for an additional 10 min. Fixed leaf tissue was then washed with sterile distilled water and frozen in liquid N_2 and stored at -80°C until needed. Prior to nucleus extraction, tissues were ground with mortar and pestle in liquid N_2 for 20 min per sample. Nuclei were isolated using the Plant Nuclei Isolation/Extraction Kit (CELLYTPN1) (Sigma-Aldrich): 3 V (volumes) NIB buffer were added to suspend the ground material and the slurry was filtered using a single layered Miracloth. The sample was spinned down for 10 min (4°C) at 1200*g*, and the supernatant was removed. The pellet was then resuspended with 0.5 mL NIBA buffer, transferred to a new tube, and 10% Triton-X 100 to a final concentration of 0.3% was added. Samples were then spinned down for 10 min (4°C) at 10,000*g*, (supernatant removed) resuspended in 0.5 mL NIBA and subjected to another round of centrifugation at 10,000*g*. The final pellet was then resuspended in 1.5 V lysis buffer (50 mM Hepes, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 10 mM sodium butyrate, 1 mM PMSF, 1× plant protease inhibitors cocktail (Sigma-Aldrich) and subdivided into 0.4-mL aliquots in 2-mL tubes for sonication using the Bioruptor Sonicator Bath (Diagnosed©). Samples were sonicated at high power level for 10 min (20 cycli of 30 s ON/OFF) with gentle mixing of the samples by invertion after 10 cycli, resulting in an average fragment size between 200-400 bp. Next, two subsequent rounds of centrifugation at 10,000*g* (4°C) for 10 min was carried out with transfer of the supernatant to new tubes in-between runs. The final extracts were subjected to pre-clearing for 2 h with 60 µL protein A magnetic Dynabeads (Thermo Fisher) per mL extract. Protein concentrations were measured using the Bradford assay, and equal protein inputs (5.5 mg) were used for all samples. ChIP reactions were performed with 5 µL anti-GFP antibody (ab290, Abcam©) incubation overnight with mild rotation at 4°C, followed by 2 h of incubation with 30 µL of protein A magnetic Dynabeads (Thermo Fisher). Washes were executed in a cold room (4°C) with two times 0.5 mL of Lysebuffer, LNDET buffer (250 mM LiCl, 1% sodium deoxycholate, 1 mM EDTA, 1% NP-40) and TE buffer (10 mM Tris, pH 8, 1 mM EDTA), in which the first 0.5 mL was used to resuspend the beads by inverting the tubes 10 times, and the second 0.5 mL was used for incubation for 5 min using mild rotation. Samples were resuspended at room temperature (RT) in 50 µL Elution buffer (1% SDS, 100 mM $NaHCO_3$, 0.2 mg/mL Proteinase K, 1 mM DTT) and eluted from the beads using a heat block at 65°C for 10 min, at 400 rpm. Supernatants were transferred to 1.5 mL DNA LoBind eppendorf tubes (Eppendorf©), and the beads were resubjected to a second round of elution using 50 µL Elution buffer and supernatants were added to the first eluates. Reverse cross-linking of the samples was then carried out overnight at 65°C on a rotation wheel. Next, samples were incubated with 1 µL of RNAse A (1 mg/mL, Qiagen) for 30 min at RT, and DNA was extracted with 32 µL EB at 65° using the PCR purification kit (Qiagen). For eluate samples, MinElute columns were used in combination, whereas standard PCR purification columns were used for input samples. Concentrations of samples were determined with Qubit™ (high dsDNA, Thermo Fischer).

Library preparation and sequencing of enriched DNA was done via Illumina sequencing on a Genome II Analyser (Illumina,

(http://www.illumina.com/applications/sequencing.html). In total, 20 million reads were carried out per sample with 125-bp long-paired-end reads. The sequenced data were processed as described (Van Leene *et al.*, 2016). Here, peak calling was done via MACS2 (Zhang *et al.*, 2008*b*) with an effective genome size of 1920Mb and MFold parameters [10, 30]. Peaks were annotated against the ZmB73_5b.60 version of the genome, and shared peaks were identified using multi-intersectBed (bedtools v2.16.1) with at least 25% overlap. All ZmAN3 samples were normalised against a merged dataset per developmental zone of the ZmNLS-GS^{yellow} control. Overlapping peaks with the full growth zone were determined using the intersect data of the full growth zone experiment samples Fb1 and Fb2 (50% overlap) compared to ZmNLS and the multi-intersect data from division zone samples D-E-F (25% overlap) using at least 25% overlap. The same procedure was carried out to determine the overlapping peaks of the full growth zone data and expansion zone multi-intersect of samples A-C-D and for expansion zone sample A separately.

REFERENCES

Adriaens ME, Prickaerts P, Chan-Seng-Yue M, van den Beucken T, Dahlmans VEH, Eijssen LM, Beck T, Wouters BG, Voncken JW, Evelo CTA. 2016. Quantitative analysis of ChIP-seq data uncovers dynamic and sustained H3K4me3 and H3K27me3 modulation in cancer cells under hypoxia. Epigenetics & Chromatin **9**, 48.

Allhoff M, Seré K, F Pires J, Zenke M, G Costa I. 2016. Differential peak calling of ChIP-seq signals with replicates with THOR. Nucleic Acid Research **44**, e153.

Baute J, Herman D, Coppens F, De Block J, Slabbinck B, Aqcua MD, Pè ME, Maere S, Nelissen H, Inzé D. 2016. Combined large-scale phenotyping and transcriptomics in maize reveals a robust growth regulatory network. Plant Physiology, pp.01883.2015.

Bommert P, Lunde C, Nardmann J, Vollbrecht E, Running M, Jackson D, Hake S, Werr W. 2005. thick tassel dwarf1 encodes a putative maize ortholog of the *Arabidopsis* CLAVATA1 leucine-rich repeat receptor-like kinase. Development **132**, 1235–1245.

Candaele J, Demuynck K, Mosoti D, Beemster GTS, Inzé D, Nelissen H. 2014. Differential methylation during maize leaf growth targets developmentally regulated genes. Plant Physiology **164**, 1350–1364.

Chen Y, Hao X, Cao J. 2014. Small auxin upregulated RNA (SAUR) gene family in maize: Identification, evolution, and its phylogenetic comparison with *Arabidopsis*, rice, and sorghum. Journal of Integrative Plant Biology **56**, 133–150.

Devailly G, Mantsoki A, Michoel T, Joshi A. 2015. Variable reproducibility in genome-scale public data: A case study using ENCODE ChIP sequencing resource. FEBS Letters **589**, 3866–3870.

DeYoung BJ, Bickle KL, Schrage KJ, Muskett P, Patel K, Clark SE. 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in *Arabidopsis*. The Plant Journal **45**, 1–16.

Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MCP. 2014. Genome-wide analysis of leafbladeless1-regulated and phased small RNAs underscores the importance of the TAS3 ta-siRNA pathway to maize development. PLoS Genetics **10**, e1004826.

Douglas RN, Wiley D, Sarkar A, Springer N, Timmermans MCP, Scanlon MJ. 2010. ragged seedling2 Encodes an ARGONAUTE7-like protein required for mediolateral expansion, but not dorsiventrality, of maize leaves. Plant Cell **22**, 1441–1451.

Feng W, Hale CJ, Over RS, Cokus SJ, Jacobsen SE, Michaels SD. 2017. Large-scale heterochromatin remodeling linked to overreplication-associated DNA damage. Proceedings of the National Academy of Sciences of the United States of America **114**, 406–411.

Gallegos JE, Rose AB. 2017. Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation. Plant Cell, tpc.00020.2017.

Goslin K, Zheng B, Serrano-Mislata A, Rae L, Ryan PT, Kwaśniewska K, Thomson B, O'Maoileidigh D, Madueno F, Wellmer F, Graciet E.. 2017. Transcription factor Interplay between LEAFY and APETALA1/ CAULIFLOWER during Floral Initiation. Plant Physiology, pp.00098.2017.

Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, Cerwick SF, Dieter JA, Duncan KE, Howard RJ, Hou Z, Löffler CM, Cooper M, Simmons CR. 2013. Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. J Exp Bot **65**, 249–260.

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.

Hoe Kim J, Tsukaya H. 2015. Regulation of plant growth and development by the GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR duo. J Exp Bot **66**, 6093–6107.

Horiguchi G, Nakayama H, Ishikawa N, Kubo M, Demura T, Fukuda H, Tsukaya H. 2010. ANGUSTIFOLIA3 Plays Roles in Adaxial/Abaxial Patterning and Growth in Leaf Morphogenesis. Plant & Cell Physiology **52**, 112–124.

Kaufmann K, Muiño JM, Østerås 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). Nature Protocols **5**, 457–472.

Lakhal W, Bader R, Stam M, Sabatel C, Poncelet D. Optimization of ChIP assay on maize tissue using Diagenode $H_3K_4me_3$, $H_3K_{27}me_3$ and H_3core antibodies

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research **22**, 1813–1831.

Li S. 2015. The *Arabidopsis* thaliana TCP transcription factors: A broadening horizon beyond development. Plant Signaling & Behavior **10**, e1044192–00.

Liang K, Keles S. 2012. Detecting differential binding of transcription factors with ChIP-seq. Bioinformatics **28**, 121–122.

Lopez-Anido C, Sun G, Koenning M, Srinivasan R, Hung HA, Emery B, Keles S, Svaren J. 2015. Differential Sox10 genomic occupancy in myelinating glia. Glia **63**, 1897–1914.

Lun ATL, Smyth GK. 2014. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. Nucleic Acids Research **42**, e95–e95.

Lun ATL, Smyth GK. 2016. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Research **44**, e45–e45.

Mao H-D, Yu L-J, Li Z-J, Yan Y, Han R, Liu H, Ma M. 2016. Genome-wide analysis of the SPL family transcription factors and their responses to abiotic stresses in maize. Plant Gene **6**, 1–12.

Morohashi K, Casas MI, Falcone Ferreyra ML, *et al.* 2012. A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. Plant Cell **24**, 2745–2764.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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..

Nolan T, Liu S, Guo H, Li L, Schnable P, Yin Y. 2017. Identification of Brassinosteroid Target Genes by Chromatin Immunoprecipitation Followed by High-Throughput Sequencing (ChIP-seq) and RNA-Sequencing. Methods Mol Biol **1564**, 63–79.

Pal B, Bouras T, Shi W, Vaillant F, Sheridan JM, Fu N, Breslin K, Jiang K, Ritchie ME, Young M, Lindeman GJ, Smyth GK, Visvader JE. 2013. Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. CellReports **3**, 411–426.

Proost S, van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K. 2015. PLAZA 3.0: an access point for plant comparative genomics. **43**, D974–81.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**, 139–140.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin SF, Palmieri C, Caldas C, Carroll JS. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature **481**, 389–393.

Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE **8**, e61005.

Shi J, Drummond BJ, Wang H, Archibald RL, Habben JE. 2016. Maize and Arabidopsis ARGOS Proteins Interact with Ethylene Receptor Signaling Complex, Supporting a Regulatory Role for ARGOS in Ethylene Signal Transduction. Plant Physiol.171(**4**): 2783–2797.

Spartz AK, Lee SH, Wenger JP, Gonzalez N, Itoh H, Inzé D, Peer WA, Murphy AS, Overvoorde PJ, Gray WM. 2012. The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. The Plant Journal **70**, 978–990.

Stark R, Brown G. 2011. DiffBind: differential binding analysis of ChIP-Seq peak data. R package version.

Takahashi H, Iwakawa H, Ishibashi N, *et al.* 2013. Meta-analyses of microarrays of *Arabidopsis* asymmetric leaves1 (as1), as2 and their modifying mutants reveal a critical role for the ETT pathway in stabilization of adaxial-abaxial patterning and cell division during leaf development. Plant Cell Physiol **54**, 418–431.

Tian B, Yang J, Brasier AR. 2011. Two-Step Cross-linking for Analysis of Protein–Chromatin Interactions. Methods in Molecular Biology. Transcriptional Regulation. New York, NY: Springer New York, 105–120.

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, Inzé 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. Nature Protocols **10**, 169–187.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L, Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. **26**, 210–229.

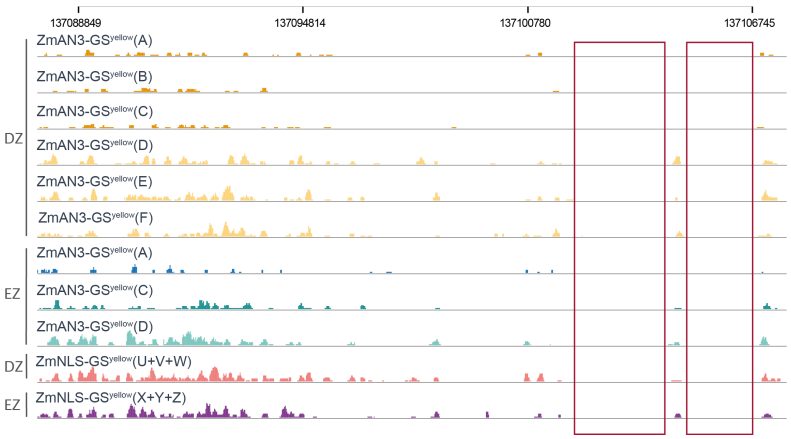
Wang L, Gu X, Xu D, Wang W, Wang H, Zeng M, Chang Z, Huang H, Cui X. 2010. miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in Arabidopsis. J Exp Bot **62**, 761–773.

Wu L, Zhang D, Xue M, Qian J, He Y, Wang S. 2014. Overexpression of the maize GRF10 , an endogenous truncated growth-regulating factor protein, leads to reduction in leaf size and plant height. Journal of Integrative Plant Biology **56**, 1053–1063.

Zhang D-F, Li B, Jia G-Q, Zhang T-F, Dai J-R, Li J-S, Wang S-C. 2008*a*. Isolation and characterization of genes encoding GRF transcription factors and GIF transcriptional co-activators in Maize (*Zea mays* L.). Plant Science **175**, 809–817.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008b. Model-based Analysis of ChIP-Seq (MACS). Genome biology **9**, R137.

SUPPLEMENTAL FIGURES AND TABLES



Supplemental Figure S1. Distribution of background peaks in the genome for all sequenced samples using IGV. Red boxes indicate read-free regions that correspond to difficult to map sequences, which is probably due to the folding status of the chromatin folding status. For the ZmNLS-GS^{yellow} control samples, reads from three replicates were merged per zone

Supplemental Table S1. Number of sequenced reads per sample

Sample		# Reads
Division zone		
ZmAN3	A	1.36E+07
	B	8.29E+06
	C	1.31E+07
	D	4.80E+07
	E	5.30E+07
	F	5.56E+07
ZmNLS	U	2.41E+07
	V	2.15E+07
	W	2.07E+07
Expansion zone		
ZmAN3	A	1.57E+07
	B	1.95E+06
	C	2.93E+07
	D	5.06E+07
ZmNLS	X	1.43E+07
	Y	2.07E+07
	Z	2.95E+07

List of reads (forward and reverse) retained per sample after clean up and removal of duplicate reads

DICUSSION AND PERSPECTIVES

GS^{YELLOW} IS A POWERFUL TOOL FOR FUNCTIONAL ANALYSIS OF PROTEIN COMPLEXES

With the vast and constant increase in publicly available whole-genome sequences generated via NGS platforms, system-wide approaches to annotate all these new sequences are becoming more and more important. To aid the study of the functions of proteins and their dynamics, we developed and evaluated a multi-functional TAP tag, GS^{yellow}, that allows the identification of protein-protein interactions through AP-MS and of protein-DNA interactions through ChIP-seq. By combining data gathered through TAP, localisation assays and ChIP-seq experiments, the previously reported function of a selected set of proteins (AN3, PPD2 and TPLATE) and their associating complexes in *Arabidopsis* or *Zea mays* was confirmed and extended (Chapter 3). Our data thus indicate the validity of the optimised ChIP-seq protocol, and also demonstrate the usability of protein tagging in combination with recombinant protein expression in functional analysis methods. There are, however, some aspects of recombinant expression of tagged proteins that have to be taken into consideration when analysing functional data.

Adding a tag might interfere with the full functionality of the protein, which could influence both the localisation and the formation of interactions of the tagged protein. Therefore, it is important to test both C- and N-terminal fusions of the bait, and to confirm that the proteins are folded properly. Validating full functionality of the fusion protein is then carried out through complementation analysis by expressing the fusion protein through its own promoter in a mutant background (Zhou *et al.*, 2011). If the transcript level of the recombinant gene is the same as its endogenous counterpart but the mutation is not complemented, this could indicate that the protein is either not folded, or that it cannot move to its locus of action, or that the tag interferes with the binding of interaction partners. Although this is a common strategy to confirm the functionality of of the fusion protein, there are two concerns regarding complementation approaches. The first one is our lack of understanding in how the exact endogenous regulatory landscape of a gene is constructed, which is especially an issue in crops like maize (Mejía-Guerra *et al.*, 2016). To circumvent the selection of a subsequence, an alternative method based on the implementation of homologous recombination for the construction of recombinant proteins was therefore tested in Chapter 4. The succesrate of this TAC recombineering technique appeared to be low, which could explain why the technology has only been used in plants in a handful of publications (Liu *et al.*, 1999; Bitrián *et al.*, 2011; Pietra *et al.*, 2013; Rips *et al.*, 2014; Han *et al.*, 2015), whereas BAC based cloning is a more popular strategy in other model species, such as mice, *E. coli* and humans. One could speculate that with the adaptation of the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 recombination mechanism, TAC-based recombineering might become of lesser interest in the near future. By transforming plant cells with a T-DNA vector that contains 1) the tag flanked by two arms that are homologous to the sides of the target site and each contain a recognition site for the Cas9, 2) a codon optimised version of the Cas9 nuclease and 3) a guide RNA (gRNA) to target the Cas9 to the genomic locus of

interest and the recognition sites in the flanking regions of the tag, any gene can be tagged endogenously. The gRNA is bound by Cas9, and will lead to double stranded DNA breaks in the genome and excision of the tag cassette. The DSB will induce HR-mediated repair at the target site, allowing integration of the tag through the homologous sequences flanking the tag. Although the system has been optimised for usage in plants (to insert a selection marker)(Baltes *et al.*, 2014; Schiml *et al.*, 2014), its implementation specifically for the insertion of a tag has not yet been reported. It is also not a certainty that all biologically interacting partners can be picked up through (T)AP-MS when using the endogenous promoter. This was examplefied with the *pAUR1:AUR1-GS*^{yellow} construct in Chapter 5. Although the fusion protein was able to complement the *aur1-2aur2-2* mutant phenotype in *Arabidopsis* seedlings, detection of the rather low abundant complexes proved to be challenging.

The second point of concern regarding complementation analyses is linked to the large amount of species for which no mutant libraries are available, or for which production of mutants is difficult, as is the case for maize and other crop species. Additionally, studying genes belonging to large gene families with redundant functions or in plants species with a high ploidy level (for instance, wheat) is also difficult, because they often fail to provide a phenotype upon mutation (Alonso and Ecker, 2006).

In our set-ups, we used mild (for maize) to strong (*Arabidopsi*s) constitutive overexpression promoters for constructs where complementation was not possible due to lack of a phenotype or mutants (as was the case for ZmAN3). When working in a WT background, overexpression is needed for competition of the bait with the endogenous counterpart. Although implementing such promoters leads to (partial) loss of developmental cues, this does not mean that dynamics can not be captured, as we showed for the ZmAN3 complex in different developmental tissues (chapter 3). However, since reverse TAPs on ZmGRF1 and ZmGRF10 showed that the first GRF preferentially binds to ZmAN3, whereas the second one was mostly associated with the other members of the ZmGIF family, ZmGIF2 and ZmGIF3 (Michiel Bontinck, personal communication), it could be that we identified both with ZmAN3 due to overexpression with the pUBIL promoter. An unbalanced situation might be created where ZmAN3 simply outcompetes the other two GIF members and co-purifies with all ZmGRFs, without reflecting the actual interactions that take place in the plant tissue. It could also be that this observed effect is linked to the insertion position of the transgene in the genome rather than to the expressionlevel. As we only performed experiments on one biological line, this possibility can not be ruled out.

PULL-DOWN VERSUS TAP: SAME SAME BUT DIFFERENT

The development of powerful statistic approaches to identify interactors of a protein using label-free quantification of MS data together with the availability of strong anti-GFP resins (Keilhauer *et al.*, 2014) and many GFP reporter lines in plants are creating an increase in pull-down experiments (Dedecker *et al.*, 2015).

Therefore, we compared the results from TAP and pull-down experiments for a set of baits (Chapter 5). Overall, TAP provided more substoichiometric interactors, which can be explained by the higher level of purity of the precipitated complex as compared to the "dirty" samples obtained with pull-down experiments. Vice versa, the pull-down experiments always provided a longer list of possible interactors, but as these interactions were not previously reported, they might indicate false positive interactors that bind to the YFP part of the tag (which are lost in TAP) or that originate from overexpression of the bait proteins in some of our set-ups. It is also possible that these interactors have only a weak or short interaction with the bait and are never picked up with TAP due to the stringent wash steps in the latter method. If the overall background levels originating from non-target tissues are reduced by using high-resolution sampling (as was done in Chapter 5 for the ZmAN3 comparison in maize with tissues from the division zone and expansion zone of the leaf), the difference between both techniques becomes less outspoken, as the reported dynamics of the complex between the different tissues could be confirmed with both (Nelissen *et al.*, 2015). Interestingly, GFP pull-downs have been reported to outperform TAP in other species, such as yeast. However, to our knowledge, direct comparison of both standard protocols on the same biological material has never been performed before. Moreover, the implemented LC-MS methodology, including the column used for LC as well as the data analysis pipeline, can have a large influence on the results of AP-MS experiments. Therefore, we decided to stick to standard protocols and found that TAP performed better than pull-down.

Although AP-MS is a strong tool to identify interacting proteins, the technique also has its challenges. First of all, MS allows identification of all interacting proteins, but cannot distinguish to which complexes they belong or which proteins directly interact with the bait. Therefore, further validation of the identified interactions has to be carried out. One approach to identify direct interactions is to use binary systems, such as Y2H or the PCA FRET/FLIM technique, where the latter can be applied to avoid expression of proteins in a yeast background. However, both techniques use a heterologous expression system that fails to provide information on the tissue context of the tested interaction (Greene *et al.*, 2015). To distinguish the different complexes associated with the bait, the purified sample can be subjected to gel purification to separate the complexes, which allows their individual detection through MS (Camacho-Carvajal *et al.*, 2004). Another option is to perform reverse TAPs, in which the identified interactors are used as bait, as was previously reported for the core complexes of AN3 (Vercruyssen *et al.*, 2014) and TPLATE (Gadeyne *et al.*, 2014).

Another point of consideration is linked to the extraction of all proteins in the sample. In general, the biological start material is a mixture of different cell types. As such, during the extraction interactions could be formed between two proteins that are normally separated from each other through compartmentalisation in the cell or through expression in different tissues. For example, when *Arabidopsis* seedlings are used, a mixture is obtained consisting of roots, leaves, SAM…, which might lead to false interactions. A possible solution for this problem is to screen the

biological material for presence of the YFP/GFP signal provided by the GS$^{yellow/green}$ tag, and only isolate signal positive tissues (*i.e.*, when the fluorescent signal is strong enough for microscopically based detection). Of course, such an approach is very laborious in small plant systems, such as *Arabidopsis*, and implies the usage of the endogenous promoter sequences. However, our interaction data obtained in the different developmental tissues of the maize leaf show that the larger organs of crop species can be subjected to this technique.

THE FUNCTION OF ZMAN3 DURING LEAF GROWTH

When the data gathered through TAP and ChIP-seq on the full growth zone are combined with the preliminary data from the differential ChIP-seq, a deeper insight into the working mechanism of ZmAN3 during leaf growth can be obtained. ChIP on the full growth zone revealed that ZmAN3 binds to the promoters of *ZmGRF3* and *ZmGRF9* (See Chapter 3, Supplemental Data S5). ChIP on the division zone specifically indicates that also the promoter of *ZmGRF11* is recognised by ZmAN3 (Chapter 6, Table 2). As TAP (and pull-down) on the division and expansion zones showed that ZmAN3 associates with ZmGRF3 and ZmGRF11 specifically in the division zone (see Chapter 3), our results indicate that the ZmAN3-ZmGRF duo might regulate its own expression through a feedback loop, which was previously also reported in *Arabidopsis* (Hoe Kim and Tsukaya, 2015).

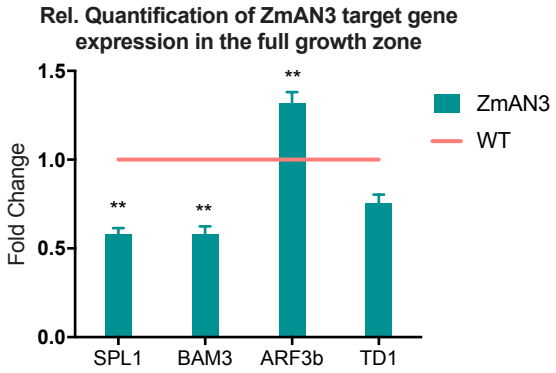


Figure 2. qPCR analysis of ZmAN3 target gene expression in the full growth zone of the pUBIL:ZmAN3-GS^{yellow} transgenic line. SPL1: SQUAMOSA PROMOTER BINDING-LIKE PROTEIN1; BAM3: BARELY ANY MERISTEM3; ARF3B: AUXIN RESPONS FACTOR3B; TD1: THICK TASSEL DWARF1. Error bars represent SE (n=2); **: significant difference with WT (*t*-test; p<0.05)

ChIP on the full growth zone and the division zone both showed that ZmAN3 binds to the intron and the promoter of *BAM3* and *TD1*, respectively. In an attempt to confirm any of the identified target genes in both ChIP experiments, we recently continued with a qPCR analysis on the full growth zone of pUBIL:ZmAN3-GS^{yellow}-expressing leaves. Interestingly, the data from this experiment showed that the expression of BAM3 is significantly lower in the full growth zone than that of WT

plants (almost half of the transcript level in WT plants, *t*-test; p = 0.024) (see Figure 2).

A similar lower transcript level is seen for TD1, although the difference to WT transcript levels was not significant (*t*-test; p-value = 0.077). In *Arabidopsis,* BAM3 has been reported to be involved in leaf shape control as well as in vascular strand development in the young leaf primordia (DeYoung *et al.*, 2005). In maize, *BAM3* expression was shown to decrease slowly along the leaf length (whereas ZmAN3 expression decreases rapidly along the leaf length), and remains present also at the leaf tip (Sekhon *et al.*, 2013), which suggests that binding of ZmAN3 could have a positive effect on the expression of BAM3. The function of TD1 in leaf development is not yet understood. However, the expression levels of TD1 in the leaf follow the exact same pattern as ZmAN3: high at the base and rapidly decreasing along the longitudinal axis of the leaf (Sekhon *et al.*, 2013), which is opposite to the qPCR results. To get a better view of the situation, an additional qPCR on the different developmental tissues of our transgenic line should be carried out. This will allow us to exclude any effect caused by the overexpression of ZmAN3 in the expanding tissue, which could provide a continuous signal is that constantly inhibits transcription of BAM3 and TD1 and thqtresults in a lower transcript level detected with qPCR for the full growth zone. Another possible target gene of ZmAN3 that plays a role in leaf development is *ZmARF3b*. Although this target was only identified with a binding site for ZmAN3 in the promoter with the differential ChIP-seq on the division zone, the qPCR analysis on the full growth zone indicated that the transcript level of ZmARF3b is enhanced in the ZmAN3-GS^{yellow} line. Congruently with this result, *ZmARF3b* was reported to have higher expression levels at the base of the leaf, with a decrease in expression along the leaf length (Sekhon *et al.*, 2013). As such, our data suggest a positive regulatory role of ZmAN3 on *ZmARF3b* expression. In *Arabidopsis,* ARF3 is a known regulator of the adaxial/abaxial identity of the leaf. Its accumulation in the abaxial side is limited by the adaxial expression of *TAS3,* a tasi-ARF encoding gene that targets ARF3 mRNA. Similar to this, *ZmARF3b* mRNA in maize is also targeted by tasiR-ARF. Moreover, maize plants that are defective in either RGD2 or LBL1, the two processors of tasiR-ARF, showed a severe defective leaf phenotype (Dotto *et al.*, 2014). In addition, mediolateral expansion of the leaf independently of the abaxial/adaxial leaf polarity was reported to be regulated through RGD2 (Douglas *et al.*, 2010). As such, our data suggest that ZmARF3b could be involved in mediolateral expansion rather than dorso-ventral polarity. However, the genome of maize also encodes four other ZmARF3 proteins. Hence, further investigation of the potential role of ZmAN3 and ZmARF3b in leaf development is needed. Finally, binding of ZmAN3 to the intron of SPL1 was found with ChIP on the full growth zone as well as with the differential ChIP-seq. SPL proteins are TFs that have a conserved role in plant growth and development. Although they have been reported to be expressed in various tissues in maize (Mao *et al.*, 2016), the specific role of the SPL1 gene is still unknown. So far, we can only suggest that ZmAN3 negatively regulates the expression of SPL1 based on the low transcript levels seen with the qPCR analysis.

Taken all of the above results together, it becomes clearer that ZmAN3 might not only be involved in the maintenance of cell division but could also influence leaf patterning during leaf growth. As our TAP experiments in *Arabidopsis* showed that AN3 also interacts with other TFs than the GRF family alone, and TAP on maize showed how well the associating complexes of AN3 are conserved between monocots and dicots, one could speculate that additional TFs might have a regulatory role through ZmAN3 on leaf development.

As a future perspective, it will be interesting to confirm any of the selected candidate target genes. Several strategies could here be followed. Using for example an inducible ZmAN3 line in combination with RNA-seq might help understand better whether ZmAN3 represses or activates the selected target genes, as it circumvents the constitutive expression of ZmAN3 in the expansion zone. The availability of a ZmAN3 mutant line will also be of great value as it can be used to identify genes that are influenced by ZmAN3 expression. One could also directly perform a ChIP on the ZmGRFs that associate with ZmAN3, because they might have a smaller amount of potential binding sites and also allow direct identification of which ZmGRF-ZmAN3 duo regulates which genes. However, given the instability of the ZmGRFs due to the miR396-mediated degradation, performing experiments on the native ZmGRFs is rather difficult. A possible solution could be the miR396 resistant ZmGRF1^{R} line that was previously reported (Nelissen *et al.*, 2015). However, by increasing the stability of ZmGRFs in the growth zone, again an artificial situation is created, as is seen by the enhanced leaf phenotype of this ZmGRF1^{R} line. As such it could be interesting to validate or screen possible target genes in a yeast background, using a Y3H system where both ZmAN3 and a ZmGRF are expressed. There is also protocol available for transient luciferase assays in protoplasts, which could also allow to circumvent the develomental expression of miR396. Another option to look into is the production of genetic crosses between the ZmGRF1^{R} or the ZmGRF10 line and the ZmAN3 mutant line. If the previously observed phenotypes dissapear in a ZmAN3 mutant background it could indicate that the initial growth enhancing effect is indeed based on the activity of ZmAN3.

It should be mentioned that for most of our experimental work, we focused on the expanding and dividing tissues. Given that almost no overlap of target genes was found between both zones, this might indicate that the regulatory role of ZmAN3 is either non-existent in the expansion zone, or that the function differs to a great extent in both tissues. As such, it might be more interesting to perform a set of functional analysis experiments (both TAP and ChIP-seq) on the zones between the expanding and dividing tissues to capture more gradient-like changes along the leaf growth zone. In addition, if results have to be compared, great care should be given into growing of the biological material, as there is always variation in plants grown in different seasons/ locations. Especially for ChIP-seq experiments, where the reproducibility is rather low, growing both the controle lines and transgenic lines should be carried out (if feasible) in parallel.

REFERENCES

Alonso JM, Ecker JR. 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nature Reviews Genetics **7**, 524–536.

Baltes NJ, Gil-Humanes J, Cermak T, Atkins PA, Voytas DF. 2014. DNA replicons for plant genome engineering. Plant Cell **26**, 151–163.

Bitrián M, Roodbarkelari F, Horváth M, Koncz C. 2011. BAC-recombineering for studying plant gene regulation: developmental control and cellular localization of SnRK1 kinase subunits. The Plant Journal **65**, 829–842.

Camacho-Carvajal MM, Wollscheid B, Aebersold R, Steimle V, Schamel WWA. 2004. Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. Molecular & Cellular Proteomics **3**, 176–182.

Costa S, Almeida A, Castro A. 2014. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: the novel Fh8 system. Front Microbiol **5,** 63

Dedecker M, Van Leene J, De Jaeger G. 2015. Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry. Current Opinion in Plant Biology **24**, 1–9.

DeYoung BJ, Bickle KL, Schrage KJ, Muskett P, Patel K, Clark SE. 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in *Arabidopsis*. The Plant Journal **45**, 1–16.

Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MCP. 2014. Genome-wide analysis of leafbladeless1-regulated and phased small RNAs underscores the importance of the TAS3 ta-siRNA pathway to maize development. PLoS Genetics **10**, e1004826.

Douglas RN, Wiley D, Sarkar A, Springer N, Timmermans MCP, Scanlon MJ. 2010. ragged seedling2 Encodes an ARGONAUTE7-like protein required for mediolateral expansion, but not dorsiventrality, of maize leaves. Plant Cell **22**, 1441–1451.

Gadeyne A, Sánchez-Rodríguez 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.

Greene CS, Krishnan A2, Wong AK3, Ricciotti E4, Zelaya RA5, Himmelstein DS6, Zhang R7, Hartmann BM8, Zaslavsky E8, Sealfon SC8, Chasman DI9, FitzGerald GA4, Dolinski K2, Grosser T4, Troyanskaya OG10. *l.* 2015. Understanding multicellular function and disease with human tissue-specific networks. Nature Genetics **47**, 569–576.

Han SW, Alonso JM, Rojas-Pierce M. 2015. REGULATOR OF BULB BIOGENESIS1 (RBB1) Is Involved in Vacuole Bulb Formation in *Arabidopsis*. PLoS ONE **10**, e0125621.

Hoe Kim J, Tsukaya H. 2015. Regulation of plant growth and development by the GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR duo. J Exp Bot **66**, 6093–6107.

Keilhauer EC, Hein MY, Mann M. 2014. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). Molecular & Cellular Proteomics **14**, 120–135.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research **22**, 1813–1831.

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. The annuals of applied statistics.

Liu YG, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D. 1999. Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proceedings of the National Academy of Sciences **96**, 6535–6540.

Mao H-D, Yu L-J, Li Z-J, Yan Y, Han R, Liu H, Ma M. 2016. Genome-wide analysis of the SPL family transcription factors and their responses to abiotic stresses in maize. Plant Gene **6**, 1–12.

Mejía-Guerra MK, Li W, Galeano NF, Vidal M, Gray J, Doseff AI, Grotewold E. 2015. Core Promoter Plasticity Between Maize Tissues and Genotypes Contrasts with Predominance of Sharp Transcription Initiation Sites. Plant Cell **27**, 3309–3320.

Nelissen H, Eeckhout D, Demuynck K, Persiau P, 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, Inzé 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.

Palmer E, Freeman T. 2004. Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays. Comparative and functional genomics **5**, 342–353.

Perduns R, Horst-Niessen I, Peterhansel C. 2015. Focus on Chromatin/Epigenetics: Photosynthetic Genes and Genes Associated with the C4 Trait in Maize Are Characterized by a Unique Class of Highly Regulated Histone Acetylation Peaks on Upstream Promoters. Plant Physiology **168**, 1378–1388.

Peret B, Swarup K, Ferguson A, Seth M, Yang Y, Dhondt S, James N, Casimiro I, Perry P, Syed A, Yang H, Reemmer J, Venison E, Howells C, Perez-Amador MA, Yun J, Alonso J, Beemster GT, Laplaze L, Murphy

A, Bennett MJ, Nielsen E, Swarup R. 2012. AUX/LAX Genes Encode a Family of Auxin Influx Transporters That Perform Distinct Functions during *Arabidopsis* Development. Plant Cell **24**, 2874–2885.

Pietra S, Gustavsson A, Kiefer C, Kalmbach L, Hörstedt P, Ikeda Y, Stepanova AN, Alonso JM. 2013. *Arabidopsis* SABRE and CLASP interact to stabilize cell division plane orientation and planar polarity. Nature Communications **4**, 2779.

Rips S, Bentley N, Jeong IS, Welch JL, Schaewen von A, Koiwa H. 2014. Multiple N-glycans cooperate in the subcellular targeting and functioning of *Arabidopsis* KORRIGAN1. Plant Cell **26**, 3792–3808.

Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. 2008. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Molecular & Cellular Proteomics **7**, 282–289.

Schiml S, Fauser F, Puchta H. 2015. The CRISPR/Cas System can be used as Nuclease for *in planta* Gene Targeting and as Paired Nickases for Directed Mutagenesis in *Arabidopsis* Resulting in Heritable Progeny. The Plant Journal **6,** 1139–1150

Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE **8**, e61005.

Venturelli S, Belz RG, Kämper A, Berger A, von Horn K, Wegner A, Böcker A, Zabulon G, Langenecker T, Kohlbacher O, Barneche F, Weigel D, Lauer UM, Bitzer M, Becker C. 2015. Plants Release Precursors of Histone Deacetylase Inhibitors to Suppress Growth of Competitors. Plant Cell **27**, 3175–3189.

Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, Jégu T, Archacki R, Van Leene J, Andriankaja M, De Bodt S, Abeel T, Coppens F, Dhondt S, De Milde L, Vermeersch M, Maleux K, Gevaert K, Jerzmanowski A, Benhamed M, Wagner D, Vandepoele K, De Jaeger G, Inzé D. 2014. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. **26**, 210–229.

Wu L, Zhang D, Xue M, Qian J, He Y, Wang S. 2014. Overexpression of the maize GRF10 , an endogenous truncated growth-regulating factor protein, leads to reduction in leaf size and plant height. Journal of Integrative Plant Biology **56**, 1053–1063.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome biology **9**, R137.

Zhou R, Benavente LM, Stepanova AN, Alonso JM. 2011. A recombineering-based gene tagging system for *Arabidopsis*. The Plant Journal **66**, 712–723.

Curriculum Vitae

Nienke Besbrugge

Fortlaan 7, 9000 Gent
☆ 20 februari 1989, Gent
✆ 0032 483 285296
✉ 206ienke.besbrugge@gmail.com

PROFESSIONAL EXPERIENCE

2012 – Current	Doctoral researcher (personal IWT grant) – Research in plant biotechnology Ghent University - VIB

EDUCATION

2017	PhD in Biochemistry and Biotechnology – Ghent University
2012	Master Biochemistry and Biotechnology – Ghent University Obtained with great distinction
2010	Bachelor Biochemistry and Biotechnology – Ghent University

SCIENTIFIC PUBLICATIONS

- **Gonzalez N, Pauwels L, Baekelandt A, De Milde L, Van Leene J, Besbrugge N, Heyndrickx KS, Cuéllar Pérez A, Durand AN, De Clercq R, Van De Slijke E, Vanden Bossche R, Eeckhout D, Gevaert K, Vandepoele K, De Jaeger G, Goossens A, Inzé D.**2015. A repressor protein complex regulates leaf growth in *Arabidopsis*. The Plant Cell **27**, 2273–2287
- **Boruc J, Deng X, Mylle E, Besbrugge N, Van Durme M, Demidov D, Vandorpe M, Eeckhout D, Nowack M, De Jaeger G, Liu B, Van Damme D**. Arabidopsis TPXL2 and TPXL3 are activators of alpha Aurora kinases and TPXL3 is essential for embryo development. In preparation.
- **Besbrugge N, Cannoot B, Eeckhout D, De Winne N, Persiau G, Van De Slijke E, Van Damme D, Van Lijsebettens M, Kulkarni S, Vandepoele K, Nelissen H, Van Leene J, De Jaeger G.** 2017. $GS^{yellow/green}$: multifunctional, "Swiss-Knife" TAP tags for functional protein analysis in monocots and dicots. In preparation for Plant Physiology.
- **Besbrugge N, Cannoot B, Eeckhout D, De Winne N, Persiau G, Van De Slijke E, Van Damme D, Van Lijsebettens M, Nelissen H, Van Leene J, De Jaeger G.** 2017/2018**.** TAP versus pull-down: Do you need two to tango? In preparation.

OTHER SCIENTIFIC COMMUNICATIONS

- **Besbrugge N, Cannoot B, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Van Leene J, Nelissen H, De Jaeger G.** 2013. Functional analysis of protein complexes in Arabidopsis and corn through multifunctional TAP-technology. Poster presentation Cold Spring Harbor Laboratories, Meeting on Plant Genomes & Biotechnology: From Genes to Networks. (Cold Spring Harbor Laboratories NY, USA, 2013)

- **Besbrugge N, Cannoot B, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Van Leene J, Nelissen H, De Jaeger G.** 2013. Characterization of growth regulatory protein complexes in plants through multifunctional TAP-technology. Poster presentation on European Plant Science Retrait. (Ghent, Belgium, 2013)
- **Besbrugge N, Cannoot B, De Winne N, Eeckhout D, Persiau G, Van De Slijke E, Van Leene J, Nelissen H, De Jaeger G.** 2013. Characterization of growth regulatory protein complexes in plants through multifunctional TAP-technology. Oral presentation on PhD Symposium Ugent Faculty of Sciences 2016. (Ghent, 2016)

TRAINING AND GUIDANCE

- Scientific writing and oral presentation (2015) Ghent. Lectured by Dr. Shirley Ellis
- Proteomics Informatics Workshop of BePA. (2016) Ghent
- Carreer Guidance (2017) Ghent. Lectured by Dr. Lucia Smit
- DynaMo Summer School on Proteomics (2013) Copenhagen, Denmark

SCIENTIFIC EXPERTISE

Protein Interaction	**Plant Genetics**	**Molecular Biology**
TAP-MS	Transformation	qPCR/RT-PCR
Pull-down	TAC recombineering	Immunoblotting
Yeast Two-Hybrid	Crossing	Molecular cloning
ChIP-seq		

QUALITIES	INTERESTS	OTHER ACHIEVEMENTS
Enthusiastic, Happy	Music + Art	Graduated summa cum laude at the Music Academy August de Boeck, Asse for Violin
Result-Oriented	Sports + Circus	Gliding: Senior Cadet in the Royal Belgian Air Cadets; holder of a gliding license
Fast-learner, Generalist	Travelling + New languages	European Driving license B since (2009)
Curious, Social	Food + Lifestyle	Erasmus Programme 1st Master, Helsinki

LANGUAGES	SOFTWARE	SCIENTIFIC SOFTWARE
Dutch	MS Office	Papers, Endnote
English	Illustrator	ImageJ
French	Finul Cut Pro	Galaxy
Spanish	Keynote, Scrivener	CLC main workbench

Supplemental Data

The Supplemental Data listed below is available for download at:

https://floppy.psb.ugent.be/public.php?service=files&t=e9ab28fab92c7bbbbbe5f84a5d4dc1da
Password: PhD Nibes

CHAPTER 3

Supplemental Data S1A. Protein details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on the *Arabidopsis* cell culture TAPs.

Supplemental Data S1B. Protein Identification details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on the seedling TAPs.

Supplemental Data S2A. List of 1535 peaks present in the 25% intersect of the PPD2-GS^{yellow} samples compared to the NLS-GS^{yellow} negative control as identified by ChIP-seq on cell cultures.

Supplemental Data S2B. List of 3124 peak sequences corresponding to 2516 genes bound by PPD2-GS^{yellow} in the first sample as identified by ChIP-seq on cell cultures using NLS-GS^{yellow} as negative control.

Supplemental Data S2C. List of 2067 peak sequences corresponding to 1690 genes bound by PPD2-GS^{yellow} in the second sample as identified by ChIP-seq on cell cultures using NLS-GS^{yellow} as negative control.

Supplemental Data S2D. List of 1095 peaks present in the 25% intersect of the PPD2-HBH samples compared to a mock negative control as identified by ChIP-seq on cell cultures.

Supplemental Data S2E. List of 2347 peak sequences corresponding to 2154 genes bound by PPD2-HBH in the first sample, as identified by ChIP-seq on cell cultures using a mock culture as negative control.

Supplemental Data S2F. List of 2046 peak sequences corresponding to 1780 genes bound by PPD2-HBH in the second sample, as identified by ChIP-seq on cell cultures using a mock culture as negative control.

Supplemental Data S3A. Gene ontology classes (Biological Process, Molecular Function, Cellular compartment) that are enriched in the list of genes bound by PPD2-GS^{yellow}, as determined by the gene ontology enrichment tool integrated in the PLAZA v3.0, Dicots (Proost *et al*., 2015) comparative genomics platform.

Supplemental Data S3B. Gene ontology classes (Biological Process, Molecular Function, Cellular compartment) that are enriched in the list of genes bound by PPD2-HBH, as determined by the gene ontology enrichment tool integrated in the PLAZA v3.0, Dicots (Proost *et al*., 2015) comparative genomics platform.

Supplemental Data S3C. Gene ontology classes (Biological Process, Molecular Function, Cellular compartment) that are enriched in the list of genes bound by PPD2-GS^{yellow} that were not found with PPD2-HBH, as determined by the gene

ontology enrichment tool integrated in the PLAZA v3.0, Dicots (Proost *et al.*, 2015) comparative genomics platform.

Supplemental Data S3D. Gene ontology classes (Biological Process, Molecular Function, Cellular compartment) that are enriched in the list of genes bound by PPD2-HBH that were not found with PPD2- GS^{yellow}, as determined by the gene ontology enrichment tool integrated in the PLAZA v3.0, Dicots (Proost *et al.*, 2015) comparative genomics platform.

Supplemental Data S4A. Proteingroups output from MaxQuant search on 5 ZmAN3-GS^{yellow} TAP experiments (5x cm1 (division zone); 5x cm4 (expansion zone)) analyzed on the Q Exactive. LFQ intensities were used for the permutation-based FDR-corrected *t*-test (Figure 4 and Supplemental Table S2) performed with the Perseus software.

Supplemental Data S4B. *Zea mays* background proteins, identified from mock TAP experiments.

Supplemental Data S4C. Additional *Zea mays* proteins that are considered as background because the *Arabidopsis* ortholog is present in *Arabidopsis* TAP background list.

Supplemental Data S5A. List of 2158 peaks corresponding to 1838 genes present in the 50% intersect of the ZmAN3-GS^{yellow} samples compared to the ZmNLS-GS^{yellow} negative control as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4.

Supplemental Data S5B. List of 2932 peak sequences corresponding to 2386 genes bound by ZmAN3-GS^{yellow} in the first sample, as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4 using ZmNLS-GS^{yellow} as negative control.

Supplemental Data S5C. List of 8977 peak sequences corresponding to 6208 genes bound by ZmAN3-GS^{yellow} in the second sample, as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4 using ZmNLS-GS^{yellow} as negative control.

Supplemental Data S5D. List of 3130 peaks corresponding to 2556 genes present in the 50% intersect of the ZmAN3-GS^{yellow} samples compared to the No Antibody (NoAB) negative control as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4.

Supplemental Data S5E. List of 3614 peak sequences corresponding to 2903 genes bound by ZmAN3-GS^{yellow} in the first sample, as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4 using No Antibody (NoAB) as negative control.

Supplemental Data S5F. List of 13472 peak sequences corresponding to 8407 genes bound by ZmAN3-GS^{yellow} in the second sample, as identified by ChIP-seq on the growth zone of *Zea mays* leaf 4 using No Antibody (NoAB) as negative control.

Supplemental Data S5G. Gene ontology classes (Biological Process, Molecular Function, Cellular compartment) that are enriched in the list of genes bound by

ZmAN3-GS^{yellow}, as determined by the gene ontology enrichment tool integrated in the PLAZA v3.0, Monocots (Proost *et al.*, 2015) comparative genomics platform. Only genes that were found in the overlap of both 50% intersects of the two ZmAN3-GS^{yellow} samples compared to ZmNLS-GS^{yellow} or the No Antibody negative control were considered.

Supplemental Data S5H. List of 393 ZmAN3 target genes in maize for which an *Arabidopsis* orthologue was found to be bound by AN3 in *Arabidopsis.* Orthologues genes were determined based on best-hit families in PLAZA Monocots v3.0 (Proost *et al.*, 2015)

CHAPTER 4

Supplemental Data S1. Protein details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on the TOR-GSrhino TAPs in *Col-0 Arabidopsis* seedlings.

CHAPTER 5

Supplemental Data S1A. Protein details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on the *Arabidopsis* cell culture TAPs.

Supplemental Data S1B. Proteingroups output from MaxQuant search on 3 SAMBA-NGS^{yellow} AP experiments Velos Orbitrap. LFQ intensities were used for the permutation-based FDR-corrected *t*-test (Supplemental Data S1C) performed with the Perseus software.

Supplemental Data S1C. Permutation-based FDR-corrected *t*-test on LFQ intensities from the pull-down on SAMBA performed with the Perseus software. LFQs of SAMBA and PSB-L (WT) samples were compared.

Supplemental data S2A. Protein details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on AUR1-GS^{yellow} TAPs in *Arabidopsis* cell cultures.

Supplemental data S2B. Protein details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) on AUR1-GS^{yellow} TAPs in *Arabidopsis* seedlings.

Supplemental data S2C. Proteingroups output from MaxQuant search on 3 AUR1 AP experiments in *Arabidopsis* seedlings. LFQ intensities were used for the permutation-based FDR-corrected *t*-test (Supplemental Figure S1 and Supplemental Data S2D) performed with the Perseus software.

Supplemental Data S2D. Significantly changing proteins between the AUR1 and WT. *t*-test analysis results from the Perseus software on a set of three independent pull-down experiments for both AUR1 and WT, analyzed by Q Exactive and relative label-free quantification (MaxLFQ) with the MaxQuant software, as presented in Supplemental Data S2C.

Supplemental data S2E. Proteingroups output from MaxQuant search on 3 AN3 AP experiments in *Arabidopsis* seedlings. LFQ intensities were used for the

permutation-based FDR-corrected *t*-test (Supplemental Figure S1 and Supplemental Data S2F) performed with the Perseus software.

Supplemental Data S2F. Significantly changing proteins between the AN3 and WT sample. *t*-test analysis results from the Perseus software on a set a set of three independent pull-down experiments for both AN3 and WT, analyzed by Orbitrap Velos and relative label-free quantification (MaxLFQ) with the MaxQuant software, as presented in Supplemental Data S2E.

Supplemental Data S3A. Proteingroups output from MaxQuant search on 6 ZmAN3-GS^{yellow} LEAF pull-down experiments in *Zea mays* (3x cm1 (division zone); 3x cm4 (expansion zone)) analyzed on Q Exactive. LFQ intensities were used for the permutation-based FDR-corrected *t*-test (Supplemental Data S3B) performed with the Perseus software.

Supplemental Data S3B. Permutation-based FDR-corrected *t*-test on LFQ intensities from the pull-down on ZmAN3 performed with the Perseus software. LFQs of cm1 (division zone) were compared between ZmAN3 and WT samples.

Supplemental Data S3C. Permutation-based FDR-corrected *t*-test on LFQ intensities from the pull-down on ZmAN3 performed with the Perseus software. LFQs of cm4 (expansion zone) were compared between ZmAN3 and WT samples.

Supplemental Data S3D. Permutation-based FDR-corrected *t*-test on LFQ intensities from the pull-down on ZmAN3 performed with the Perseus software. LFQs of ZmAN3 specific cm1 (division zone) and cm4 (expansion zone) samples were compared.

CHAPTER 6

Supplemental Data S1A. Peak data for all analyzed samples in the ChIP-seq experiments on ZmAN3-GS^{yellow} in the division and expansion zone of *Zea mays* leaves as compared to the ZmNLS-GS^{yellow} negative controls.

Supplemental Data S1B. MACS2 and Intersect data from all ChIP-seq experiments carried out on ZmAN3-GS^{yellow} in the division zone and expansion zone of *Zea mays* as compared to the merged ZmNLS-GS^{yellow} zone specific controls.