UNIVERSITY OF CALIFORNIA SAN DIEGO

Role of Chromatin and Transcriptional Regulation in the DNA Damage Response

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Neeraja Vegesna

Committee in charge:

Professor Julie A. Law, Chair
Professor Joanne Chory
Professor Arshad Desai
Professor Amy E. Pasquinelli
Professor Lorraine Pillus

2021

The Dissertation of Neeraja Vegesna is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents, Ravi Raju and Santha Kumari
And my brother, Tarun.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

'	minutes
AGRIS	Arabidopsis gene regulatory information server
BER	Base excision repair
BRAT	Busch-lab Root Analysis Toolchain
cDNA	Complementary DNA
ChIP-seq	Chromatin immunoprecipitation sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAP-seq	DNA affinity purification sequencing
DDR	DNA damage response
DE	Differentially expressed
diRNA	DSB-induced small RNA
DNA	Deoxyribonucleic acid
DREM	Dynamic Regulatory Events Miner
DSB	Double strand breaks
EMS	Ethyl methanesulfonate
FC	Fold change
FDR	False discovery rate
FISH	Fluorescence In Situ Hybridization
G2/M	Gap 2/ Mitosis (in cell cycle)
GFP	Green fluorescent protein
GO	Gene ontology

Gy	gray (unit of gamma irradiation)
h	Hours
HR	Homologous recombination
IP/MS	Immunoprecipitation mass spectrometry
kb	Kilobase
KD	Knock down
lncRNA	Long non-coding RNA
LOF	Loss of function
LS½	Half strength Linsmaier and Skoog Medium
Mb	Megabase
me	methylation
mg	Milligram
min	minutes
miRNA	MicroRNA
ml	Millilitre
mRNA-seq	messenger RNA sequencing
MS½	Half strength Murashige and Skoog medium
ncRNA	Non-coding RNA
NER	Nucleotide excision repair
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
p	Phosphorylation
PANTHER	Protein ANalysis THrough Evolutionary Relationships

PCD	programmed cell death
PCR	Polymerase chain reaction
PICh	Proteomics of isolated chromatin
PTM	Post-translational modifications
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
SDM	Site directed mutagenesis
SDSA	Synthesis dependent strand annealing
SHR	somatic homologous recombination
siRNA	Small interfering RNA
SMART	Simple Modular Architecture Research Tool
sRNA	Small RNA
SSA	Single strand annealing
SSB	Single strand breaks
SSBR	Single strand break repair
su	Sumoylation
T-DNA	Transfer DNA
TF	Transcription factor
ub	Ubiquitylation
UTR	Untranslated region
UV	Ultraviolet
γ	Gamma
γ-IR	Gamma irradiation

µg	micro gram
µM	Micro molar

LIST OF FIGURES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDGEMENTS

I would like to first thank my mentor and guide, Professor Julie A. Law. Her passion for science and optimistic outlook have fostered a positive environment for me to grow as a scientist. Her persistence, leadership by example and meticulousness have been aspirational and inspired me every day. I have learned a great deal about how to conduct science from her and will forever be grateful.

I'm lucky to have had amazing colleagues and mentors in the Law lab. I would like to thank Clara Bourbousse, Ana Marie Palanca, Ceyda Coruh, Elisa Iniesto Sanchez, Ming Zhou, Mina Rostamza, Matt Simenc, Guanghui Xu and Laura Martins for their assistance at various stages of my research, for a cherished time spent together in the lab, and in social settings! I am extremely grateful to Ceyda and Ana Marie, who have also become great friends, and played a continuous role in ensuring that I completed the PhD successfully. I'm privileged to have worked with excellent undergraduate volunteers: Katherine Chang, Amy Eppert, Nicole Talaba and Elaine Ha who have taught me a great deal about being a mentor.

I am fortunate to be a part of Plant Biology at Salk Institute for Biological Studies. I would like to extend my profound gratitude to Professor Joanne Chory, Professor Joe Ecker, Professor Wolfgang Busch and their labs for sharing their resources and providing constructive suggestions on my research. I would like to thank Salk Institute for Biological Studies for various opportunities, including, fellowships and awards such as the Salk Women in Science Award, Jesse and Caryl Philips Foundation Fellowship and Chapman Charitable Trust Graduate Student Scholar Award.

It was a great experience to be a part of the UCSD/Salk Biology PhD program. The support I received from UCSD Division of Biological Sciences (specifically, Melody Bazyar, Suzi Harlow,

Cathy Pugh, Marifel Alfaro, Donna Aizuss, Lien Ngo, Samantha Crook), CAPS and ISPO was crucial for the successful completion of my Ph.D. I am grateful for the opportunity to serve as a Graduate Teaching Mentor and all the students I interacted with, which taught me a lot about teaching and communication. I am also thankful to the Introduction to College Teaching course which opened my mind to new perspectives.

I would like to thank my committee members, Professor Joanne Chory, Professor Arshad Desai, Professor Amy E. Pasquinelli and Professor Lorraine Pillus for providing me with crucial pieces of wisdom and keeping me on track during the Ph.D.

I would like to thank all the teachers and professors, that educated me and helped me get here. Specifically, I would like to extend my gratitude to my Math teacher, Vijaya Lakshmi, who set me up with a strong STEM foundation; professors at the Department of Biotechnology and Biochemical engineering in Indian Institute of Technology, Kharagpur; and my mentors at Energy Biosciences Institute (UCB), Harshal Chokhawala who got me interested in Molecular biology, which motivated me to pursue a Ph.D.

I am truly blessed to have met a wonderful group of friends who made San Diego home. I would like to thank Ullas, Sid, Mrittika, Prit, Nandu, Akshat, Annie, Kanika, Aekaansh, Sunandha, Vignesh, Raghu, Swetha, Sourish, Tapan, Tandrilla, Animesh, Sumitash, Aravind, Mihir, Bharat and Nuttida for being a part of, and supporting my journey at various times. I am especially grateful to Prit, Nandu and Tapan, whose friendships I deeply cherish and for getting me through some incredibly difficult times. I am incredibly grateful to my housemates, David and Genevieve for their kindness, thoughtfulness and getting me through 2020. I would also like to thank my friends from undergrad, Anindita, Spandana, Anusha, Spurthi, Yamini, Soumya, Debadrita and Neha for their everlasting friendships.

All of this would never have been possible without the incredible support of my family. I would like to express my utmost gratitude and heartfelt appreciation to my parents, Santha Kumari Vegesna and Ravi Raju Vegesna; and my brother, Tarun Vegesna for their unconditional support through this arduous journey. Specifically, I would like to thank my mom for always being there, her progressive outlook and shielding me from societal pressures; my dad for supporting my aspirations and setting an exceptional example of persistence, resilience, and success; and my brother for his wisdom, beyond his age. I would like to thank all my extended family, specifically my aunt Prabhavati and uncle Suryanarayana Raju. I would also like to thank my cousin, Prudhvi for looking out for me, as I navigated my life in the US.

Chapter 2 is made up of a reprint of a published manuscript. The published manuscript is: Bourbousse, Clara; Vegesna, Neeraja and Law, Julie A. "**SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis***." Proceedings of the National Academy of Sciences 115.52 (2018): E12453-E12462. The dissertation author is the secondary author of this paper.

Chapter 3, in part, is made up of a manuscript in preparation. The manuscript in preparation is: Neeraja Vegesna, Clara Bourbousse, Yasaman Jami-Alahmadi, Ana Marie S. Palanca, James A. Wohlschlegel, Julie A. Law**. "Two *Arabidopsis* YEATS domain proteins facilitate DNA damage repair via their histone binding domains"** *(in preparation)*. The dissertation author will be the first author of this manuscript in preparation.

VITA

EDUCATION

2021 Doctor of Philosophy, University of California San Diego, USA

2014 Master of Technology (Honours), Biotechnology and Biochemical Engineering, Indian institute of Technology Kharagpur, India

2014 Bachelor of Technology (Honours), Biotechnology and Biochemical Engineering, Indian institute of Technology Kharagpur, India

FELLOWSHIPS & AWARDS

- Salk Women in Science Award (2018), Salk Institute for Biological Studies.
- Jesse and Caryl Philips Foundation Fellowship (2018), Salk Institute for Biological Studies.
- Chapman Charitable Trust Graduate Student Scholar Award (2015-2018), Salk Institute for Biological Studies.
- Institute Silver Medal (2014), Indian institute of Technology Kharagpur, India.
- J.C. Ghosh Memorial Endowment Prize (2013), Indian institute of Technology Kharagpur, India.

PUBLICATIONS

- C. Bourbousse, **N. Vegesna**, and J A. Law. "SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis*." Proceedings of the National Academy of Sciences (2018).
- G. Bhowmick, **N. Vegesna**, and R. Sen. "Process design for augmentation and spectrofluorometric quantification of neutral lipid by judicious doping of pathway intermediate in the culture of marine *Chlorella variabilis* for biodiesel application." *Bioresource technology* (2015).

- H A. Chokhawala, C M. Roche, T W. Kim, M E. Atreya, **N. Vegesna**, C M. Dana, H W. Blanch, and D S. Clark. "Mutagenesis of *Trichoderma reesei* endoglucanase I: impact of expression host on activity and stability at elevated temperatures." *BMC biotechnology* (2015).

PUBLICATIONS IN PREPARATION

- N. Vegesna, C. Bourbousse, Y. Jami Alahmadi, A.M.S. Palanca, J. Wohlschlegel, J.A. Law. "Two *Arabidopsis* YEATS domain proteins facilitate DNA damage repair via their histone binding domains". (*in preparation*)

TEACHING & LEADERSHIP

- Served as Graduate Teaching Mentor, UCSD (2018-2019)
- Organized BGSE205 Gene Expression section (2016-2017)

ABSTRACT OF THE DISSERTATION

Role of Chromatin and Transcriptional Regulation in the DNA Damage Response

by

Neeraja Vegesna

Doctor of Philosophy in Biology

University of California San Diego, 2021

Professor Julie A. Law, Chair

DNA is the repository of genetic information and thus, it is important to maintain its integrity over time and generations. However, DNA can be damaged by a multitude of endogenous and exogenous damaging agents leading to genome instability. In order to maintain genomic integrity and organismal fitness, organisms have evolved highly conserved mechanisms – collectively termed the DNA damage response (DDR) – that coordinate key biological processes

needed to repair damaged DNA, including cell cycle arrest, gene regulation, DNA repair and programmed cell death. As a first step, the DDR includes the recognition of the DNA lesion in the context of the local chromatin landscape followed by activation of signaling cascades and transcriptional programs, depending on the type of DNA damage. Two critical, but as yet poorly understood aspects of the DDR in the plant model, *Arabidopsis thaliana*, are: (**1**) the kinetics and regulatory networks controlling the expression of genes orchestrating key biological processes during the DDR and (**2**) the roles of specific chromatin modifications and chromatin modifying complexes in regulating the process of DNA repair. In the studies comprising my thesis, we worked to shed light on both these knowledge gaps using a combination of genetic, genomic, and biochemical approaches. **Chapter one** introduces what is currently known about the transcriptional and chromatin-mediated aspects of the DNA damage response, specially focusing on plants. **Chapter two** presents the development of a temporal model of the *Arabidopsis* transcriptional response to DNA damage as well as the identification and characterization of key factors that regulate this transcriptional network. **Chapter three** outlines the setup of a reverse genetic screen to identify candidate chromatin-associated factors required for DNA repair. It also includes the characterization of a candidate chromatin reader, YAF9B, and its homolog YAF9A, in DNA repair. **Chapter four** concludes by synthesizing the core achievements of the dissertation and suggesting future directions. Together, these chapters have helped understand the multifaceted roles of chromatin in orchestrating DNA repair by revealing the (1) dynamics of transcriptional regulation during the DNA damage response and (2) demonstrating roles for two histone reader proteins, YAF9B and YAF9A, in double strand break repair via homologous recombination-like mechanisms.

Chapter 1 THE DNA DAMAGE RESPONSE – AN INTRODUCTION

CHAPTER SUMMARY

Organisms are subjected to numerous endogenous and exogenous stressors which contribute to DNA damage. Since DNA encodes all the instructions used to regulate the growth, development, and functions of said organism, it is critical to maintain DNA integrity and stability throughout the lifespan of the organism as well as between generations. Although mechanisms underlying various DNA damage processes are rapidly emerging, the role of transcriptional regulation and chromatin remains relatively unexplored in the *Arabidopsis* DNA damage response (DDR). This chapter is an overview of the understanding of the DDR prior to our studies that highlights key gaps in knowledge.

THE DNA DAMAGE RESPONSE PATHWAYS

Living organisms are constantly exposed to harmful agents that arise from either exogenous sources (*e.g.* UV light, radiation, chemical exposure, etc.) or endogenous cellular processes (*e.g.* metabolism, replication errors, etc.) and result in different kinds of DNA damage, such as strand crosslinking, mismatches, base damage, as well as single and double strand breaks (SSBs and DSBs, respectively)(Helleday, Eshtad, & Nik-Zainal, 2014). When damaged, DNA must be repaired in order to prevent negative outcomes such as genome instability. To this end, cells orchestrate highly conserved DNA damage response (DDR) pathways, which consists of both transcriptional and post-translational changes and can vary in their nature depending on the type of DNA damage, the tissue that is damaged, and even the stage of the cell cycle (Culligan, Robertson, Foreman, Doerner, & Britt, 2006; Z. Wang, Schwacke, & Kunze, 2016). For instance, mismatched DNA bases are corrected by removal via base excision repair (BER); pyrimidine dimers are corrected by nucleotide excision repair (NER); crosslinked DNA are repaired via inter-strand crosslink repair(Hoeijmakers, 2009; Moldovan & D'Andrea, 2009); SSBs are repaired by single-strand break repair (SSBR)(Caldecott, 2008); whereas the DSBs are processed either by non-homologous end joining (NHEJ) or homologous recombination (HR)(Scully, Panday, Elango, & Willis, 2019). NHEJ resolves DSBs quickly by ligating the broken ends but is error prone. HR precisely restores the genomic sequence by utilizing a homologous DNA template and is more common in mitotic cells in meristematic tissues, where high fidelity repair is needed. Of all the different kinds of DNA damage, DSBs are particularly dangerous as improper repair might lead to chromosomal deletions or translocations, resulting in genetic maladies such as cancer in mammals and developmental defects in plants. Unlike other damages such as SSBs, DSBs lack an undamaged complementary strand that could be used as a template to repair the damage, making

DSB repair highly complex(Aguilera & Garcia-Muse, 2013; Rodgers & McVey, 2016). The response to DSBs involves DNA damage sensing by sensor proteins, signal transduction by transducer proteins, and finally the activation of effector proteins that repair DSBs or regulate transcriptional programs. The transcriptional response to DSBs controls downstream processes critical for DNA repair and genome stability, including the activation of cell cycle checkpoints to avoid transcription at damage sites, the expression of genes involved in DNA repair, and, when the damage is severe enough, the initiation of programmed cell death(Einset & Collins, 2015; K. O. Yoshiyama, Kobayashi, et al., 2013). Given the catastrophic consequences of unrepaired or improperly repaired DSBs, DSB repair pathways have been intensely investigated across multiple organisms, especially mammals from the cancer perspective. The study of DSB repair in plants has uncovered both conserved and plant-specific aspects of repair, which are compared with mammalian insights and summarized below.

DSB damage signaling in *Arabidopsis*

The DNA damage response in yeast, plants, and mammals is conserved on many levels(Amiard, Gallego, & White, 2013; T. T. Su, 2006; K. O. Yoshiyama, Sakaguchi, & Kimura, 2013). For instance, DSBs in all these organisms are sensed by the MRN (MRE11, RAD50, and NBS1) and Ku70/80 complexes, which in turn activate the serine/threonine kinases, ATAXIA-TELANGIECTASIA MUTATED (ATM) and ATAXIA-TELANGIECTASIA MUTATED AND RAD3-RELATED (ATR). These kinases respond to different types of DNA damage and then initiate signaling cascades through the posttranslational modification of target proteins(Bleuyard, Gallego, & White, 2006; Ciccia & Elledge, 2010; Roitinger et al., 2015; Shiloh & Ziv, 2013; K. O. Yoshiyama, Kobayashi, et al., 2013). In line with such important DDR functions, *atm* and *atr*

mutants are viable in plants, but are hypersensitive to DSBs, including those induced by gamma (γ)-irradiation(Culligan et al., 2006). Recent proteomic studies in plants demonstrated that ATM and ATR phosphorylate serine-glutamine/threonine-glutamine (SQ/TQ) motifs on target proteins, revealing both conserved and organism-specific factors(Ciccia & Elledge, 2010; Knoll, Fauser, & Puchta, 2014; T. T. Su, 2006; K. O. Yoshiyama, Sakaguchi, et al., 2013). For instance, the ATM/ATR mediated phosphorylation histone variant H2A.X is conserved in both mammals and plants. On the other hand, several well characterized targets of mammalian ATM/ATR, including the CHK1/CHK2 transducer proteins and the p53 transcriptional effectors(Bartek, Falck, & Lukas, 2001; Y. Chen & Sanchez, 2004) have not been identified in plants. In mammals, p53 is a critical transcription factor (TF) that governs the expression of many target genes involved in cell cycle control, DNA repair, apoptosis and senescence in response to DNA damage and is dubbed as the tumor suppressor protein(Rozan & El-Deiry, 2007). The absence of important factors such as p53 demonstrates that at least some of the transducers and effectors of the plant DDR are not conserved and that the identification and characterization of such plant-specific effectors is necessary to understand how the DNA damage response is coordinated in plants.

Effectors: Role of SOG1 in mediating plant DNA damage response

Despite plant genomes lacking a *p53* homolog there is a functional equivalent that plays important roles in the plant DDR. This factor, named *Suppressor Of Gamma-response 1* (*SOG1*), is a transcription factor of the NAC (NAM, ATAF1/2, and CUC2) family that was isolated through a genetic screen as a suppressor of the γ-irradiation induced growth arrest observed in the *uvh1* (*UV-hypersensitive 1*) mutant(Preuss & Britt, 2003). *Arabidopsis sog1* mutant plants exposed to DNA damaging agents display defects in gene regulation(K. Yoshiyama, Conklin, Huefner, &

Britt, 2009), cell cycle arrest(Preuss & Britt, 2003), programmed cell death(Furukawa et al., 2010), endoreduplication(Adachi et al., 2011), DNA repair, and genome stability(Preuss & Britt, 2003; K. Yoshiyama et al., 2009), demonstrating that SOG1 is critical for the plant response to DNA damage. These DDR defects in *sog1* mutants, and the regulation of SOG1 in an ATM and ATR-dependent manner via phosphorylation of conserved SQ motifs(K. O. Yoshiyama, Kaminoyama, Sakamoto, & Kimura, 2017; K. O. Yoshiyama, Kobayashi, et al., 2013), have led to SOG1 being functionally equated with p53, and dubbed the master regulator of the plant DDR(K. O. Yoshiyama, 2016; K. O. Yoshiyama, Sakaguchi, et al., 2013).

Although SOG1 has been shown to influence key DDR processes(Adachi et al., 2011; Biever, Brinkman, & Gardner, 2014; P. Chen & Umeda, 2015; Furukawa et al., 2010; Hu, Cools, Kalhorzadeh, Heyman, & De Veylder, 2015; Pedroza-Garcia et al., 2017; Preuss & Britt, 2003; Sjogren, Bolaris, & Larsen, 2015; Yi et al., 2014; K. Yoshiyama et al., 2009), insights about its roles in gene regulation were based on transcriptional profiling experiments performed at single timepoints after exposure to DSB inducing stressors [gamma irradiation (γ-IR) after 2 hours(K. Yoshiyama et al., 2009) or zeocin after 1.5 hours(Ogita et al., 2018)]. However, the DDR spans a much broader time frame. For instance, the marking of the chromatin surrounding DSBs with γH2A.X occurs mere minutes after damage(Charbonnel, Allain, Gallego, & White, 2011; Charbonnel, Gallego, & White, 2010; J. D. Friesner, Liu, Culligan, & Britt, 2005). On the other hand, DDR events such as cell cycle regulation requires hours to days(Charbonnel et al., 2010; Preuss & Britt, 2003). As these early and late DDR events occur outside of the transcriptionally profiled time span, expanding our understanding of the transcripitoinal response to DNA damage to capture a larger time frame will provide insights into how the activities of SOG1 help orcheestrate the DDR and ensure genome stability.

Another unknown is how SOG1 directly controls the DDR transcriptional network. SOG1, like other NAC TFs, contains a β-fold DNA binding domain and transcription regulatory domain. The transcription regulatory domain contains 5SQ motifs that are phosphorylated by ATM/ATR and are required for the full activation and DNA binding activity of SOG1(K. O. Yoshiyama et al., 2017). Various NAC transcription factors have been involved in regulating transcriptional cascades required for growth, development, biotic, and abiotic stress response(Lindemose et al., 2014; Mohanta et al., 2020; Shao, Wang, & Tang, 2015). In the DDR, only a few SOG1 targets have been identified using ChIP-qPCR(P. Chen & Umeda, 2015; Sjogren et al., 2015; Weimer et al., 2016; Yi et al., 2014), leaving it unclear how much of the DDR is directly regulated by SOG1. A next generation sequencing (NGS)-based, comprehensive identification of SOG1 downstream targets is required to understand what part of the transcriptional response is directly targeted by SOG1 and how it orchestrates the transcriptional networks controlling downstream DDR processes.

Regulation of DDR processes in *Arabidopsis*

The DDR is a fundamental process to tolerate DNA damage, maintain genomic integrity, and ensure the faithful transmission of genetic information from one generation to the next. While some of the regulatory mechanisms required for these processes are conserved across eukaryotes, plants have evolved novel regulators of the DDR, such as SOG1. This is possibly due to a requirement for an efficient and sensitive DDR system because of (1) their constant exposure to DNA damage as a result of their sessile nature(Hu, Cools, & De Veylder, 2016) and (2) post-embryonic plant development, which is achieved by the activity of meristematic cells that divide throughout the plant's life(Beck, 2010; Jurgens, 2003). While SOG1, and some of its target genes,

have been shown to be crucial for the execution of subsets of the DDR, a comprehensive study of the transcriptional response in response to DNA damage, is yet to be performed. The current understanding of how the DDR transcriptional regulation facilitates key downstream processes important for DNA repair and genome stability, as well as major gaps in our understanding of the regulation of these processes, are summarized below.

DNA repair factors. Various DNA repair factors, responding to different types of DNA damage, have been identified from fungi to animals(Friedberg, Walker, Siede, & Wood, 2005; Wood, 2020). Components of these DNA repair pathways are mostly (but not completely), conserved between animals and plants(Kimura & Sakaguchi, 2006), suggesting that animals and plants have largely similar DNA repair machinery. Based on homology, various *Arabidopsis* HR repair factors (such as BRCA1, BRCC36A(Block-Schmidt, Dukowic-Schulze, Wanieck, Reidt, & Puchta, 2011), INO80(Fritsch, Benvenuto, Bowler, Molinier, & Hohn, 2004), RAD51(Doutriaux, Couteau, Bergounioux, & White, 1998; Y. Wang et al., 2014), RAD17(Heitzeberg et al., 2004), RAD5A(Mannuss et al., 2010), RecQ14A(Bagherieh-Najjar, de Vries, Hille, & Dijkwel, 2005), AGO2(Gao et al., 2014; Wei et al., 2012), MMS21(Yuan et al., 2014)) and NHEJ repair factors (such as Ku70/Ku80(J. Friesner & Britt, 2003; Tamura, Adachi, Chiba, Oguchi, & Takahashi, 2002), XRCC4(West, Waterworth, Jiang, & Bray, 2000), LIG4(van Attikum et al., 2003), PARP1, PARP2(Jia, den Dulk-Ras, Shen, Hooykaas, & de Pater, 2013)) have been studied. Characterization of some key DNA repair factors (e.g., BRCA1, RAD17, and PARP1) have shown that they are strongly induced after DNA damage in *Arabidopsis*(Gu et al., 2019; Heitzeberg et al., 2004; Sjogren et al., 2015) in a SOG1-dependent manner, while their mammalian counterparts are constitutively expressed(Papatheodorou et al., 2020; Rieger & Chu, 2004) and/or only weakly induced after DNA damage(Ding et al., 2013; Papatheodorou et al., 2020). Therefore, despite

having some conserved machinery, this observation suggests non-conserved gene regulatory mechanisms controlling DNA repair across species. Additionally, while a majority of the plant DNA repair factors have been shown to be transcriptionally induced by SOG1, the transcriptional response to DNA damage in animals is not exclusively regulated by p53(K. O. Yoshiyama, Kimura, Maki, Britt, & Umeda, 2014; K. O. Yoshiyama, Sakaguchi, et al., 2013). Given these observations, further identification and characterization of genes that are targeted by SOG1 would enable further understanding of the DDR in plants.

Cell cycle regulation. DNA damage leads to cell cycle arrest to allow DNA repair to occur before DNA replication or mitosis begins, making it crucial to maintain genome integrity. In mammals, this process is coordinated by cell-cycle specific kinases and their inhibitors(Ciccia & Elledge, 2010; Harper & Elledge, 2007; Hu et al., 2016). In *Arabidopsis*, it was shown that SOG1 is responsible for upregulating the G2-specific *CYCLINB1(*Schnittger & De Veylder, 2018), the plant-specific CDK inhibitors *SMR5* and *SMR7*(Weimer et al., 2016; Yi et al., 2014) and *WEE1*(Cools & De Veylder, 2009; De Schutter et al., 2007), while also being important for downregulating both *CDKs* and *KNOLLE* cell cycle factors(Missirian, Conklin, Culligan, Huefner, & Britt, 2014; K. Yoshiyama et al., 2009). However, as previously mentioned, DDR-associated cell-cycle responses occur in a timeframe spanning hours to days and the available transcriptional profiling (at 1.5h and 2h)(Ogita et al., 2018; K. Yoshiyama et al., 2009) are not able to capture the whole spectrum of cell cycle regulation. Therefore, the mechanism and temporal regulation by which SOG1 regulates the cell cycle arrest is yet to be deciphered.

Programmed Cell Death (PCD). PCD eliminates severely damaged cells, a mechanism that reduces the risk of accumulating cells with a compromised genome. In plants, PCD predominantly occurs in stem cells upon DNA damage, as they are the progenitors of all somatic and reproductive

cells(Dijkwel & Lai, 2019). In mammals, activation of the cell death program upon severe DNA damage is governed by the ATM- and ATR-dependent activation of p53 and mediated by its downstream targets (apoptotic proteins like PUMA and NOXA)(L. N. Zhang, Li, & Xu, 2013). In *Arabidopsis*, it has been shown that the ATM and ATR kinases(Fulcher & Sablowski, 2009) and SOG1(Furukawa et al., 2010; Johnson et al., 2018) are required for cell death in stem cells following DNA damage. However, *Arabidopsis* equivalents for PUMA and NOXA were not identified, suggesting further work is required to identify SOG1-affected, plant-specific cell death factors. In addition, accumulation of dead cells in *Arabidopsis* was shown using propidium iodide staining in roots 24h after DNA damage(Fulcher & Sablowski, 2009), a time at which transcriptional profiles after damage are unknown. Therefore, transcriptionally-induced factors responsible for DNA damage–induced programmed cell death are yet to be identified and characterized.

Endoreduplication. Endoreduplication is a common process in which DNA is replicated without cell division, resulting in elevated genomic DNA content and polyploidy(Shu, Row, & Deng, 2018). Apart from cell cycle arrest and PCD, plant DNA stress checkpoints can activate endoreduplication as a third response to DSB stress(Adachi et al., 2011). Upon DNA damage, endoreplication accounts for continued growth in the absence of cell division and these endoreduplicating cells rarely resume cell division, thereby preventing the transmission of DNA errors across tissue or offspring(Fox & Duronio, 2013; Lee, Davidson, & Duronio, 2009; Shu et al., 2018). Upon detection of DSBs, ATM and ATR were shown to play redundant roles in controlling endoreduplication in a SOG1-dependent manner, wherein SOG1 controls the repression of G2/M-specific cell cycle genes (including mitosis-specific cyclins) and the activation of cell cycle–inhibitory genes such as *SIM*, *SMR1*, *SMR5*, and *WEE1*(Adachi et al., 2011). In

addition, several TFs (*TCP15*(Li, Li, & Dong, 2012), *ARR2*(Takahashi & Umeda, 2014), etc.) and genes (*CCS52A1*(Breuer et al., 2012; Heyman, Polyn, Eekhout, & De Veylder, 2017), *DEL1*(Vlieghe et al., 2005), etc.) have been shown to control the endoreduplication in plant development (non-DNA damage conditions). Perhaps these genes might be directly or indirectly regulated by SOG1 after DNA damage, revealing insights about shared components of the damage-associated and development-associated endoreduplication pathways.

Control of Downstream Transcription Factors. The recognition of DSBs initiates a massive transcriptional response in an ATM- and ATR-mediated process that requires the SOG1 transcription factor(K. Yoshiyama et al., 2009). Of all the genes regulated as part of this response, only two genes were known to be directly targeted and induced by SOG1(Yi et al., 2014). However, what fraction of the transcriptional response is directly targeted by SOG1, if SOG1 is an activator or a repressor, and how the genes targeted by SOG1 influence specific DDR processes are yet to be determined. Moreover, how and to what extent TFs targeted by SOG1 propagate the massive DDR transcriptional cascade is unknown. In addition, despite this central role of SOG1, recent studies have revealed SOG1-independent pathways in the plant DDR. Firstly, RBR1 mediates repression of several DDR genes in an E2Fa-dependent manner(Biedermann et al., 2017; Horvath et al., 2017). Secondly, SNI1 mediates expression of various defense and broad immune response genes required for systemic resistance and is important for proper homologous recombination(Durrant, Wang, & Dong, 2007; S. Wang, Durrant, Song, Spivey, & Dong, 2010). Therefore, further temporal characterization of the transcriptional response and identifying the direct targets if SOG1 will help understand the SOG1-dependent and -independent transcriptional cascades responsible for the dynamics and regulation of the DDR processes.

DSB REPAIR IN THE CONTEXT OF CHROMATIN

Effect of chromatin landscape on DNA repair

In the nucleus, DNA exists in the form of chromatin, where 146 bp of DNA is wrapped around a core histone octamer (two each of H2A, H2B, H3, and H4) making up the nucleosome and histone H1 binds to the linker DNA between nucleosomes(K. Zhou, Gaullier, & Luger, 2019). The chromatin landscape is dictated by (1) the exchange of canonical histones with variant forms that play distinct and specialized roles(Buschbeck & Hake, 2017); (2) post-translational modifications of histones (acetylation, methylation, phosphorylation, ubiquitylation, SUMOylation)(Miller & Jackson, 2012) and DNA (cytosine) methylation(Law & Jacobsen, 2009); (3) chromatin remodeling activities to reposition nucleosomes by sliding, evicting, or depositing these proteins in the genome(Sundaramoorthy & Owen-Hughes, 2020); and (4) the involvement of various non-coding RNAs(Waititu, Zhang, Liu, & Wang, 2020; Wu, Liu, Qi, Cai, & Xu, 2020). The DNA-histone complex is mediated by the electrostatic interactions between negatively charged DNA and positively charged histones. Thus, modifications of amino acid residues have the potential to affect the strength of the associations between DNA and histones, which in turn can disrupt or stabilize DNA/protein and protein/protein interactions(Bannister & Kouzarides, 2011). These modified residues can also create binding sites for chromatin "reader" proteins, which in turn recruit and facilitate further chromatin modifications by chromatin "writer" proteins (that catalyze the deposition of new modifications) or "eraser" proteins (that remove existing modifications). Crosstalk between multiple chromatin-modifying complexes and their targets result in a diverse chromatin landscape, which has profound influences on all DNA processes including DNA repair. In mammals, several histone variants and chromatin modifications have been shown to be important for DNA repair and the readers, writers, and erasers that interact with

these proteins are emerging. However, in plants, research exploring the role of chromatin in DNA repair is lagging and the reader, writer, and eraser proteins important for DNA repair are almost completely unknown. Characterization of such factors across organisms would shed light on the order of chromatin-associated events mediated by various modifications after damage and provide insights into how chromatin modulation regulates proper repair.

Chromatin modifications in DNA repair

In mammals, several key histone modifications and histone variants have been reported to participate in DSB repair via various mechanisms including signaling DNA damage, allowing repair conducive open chromatin states, and influencing repair pathway choice (J. J. Kim, Lee, & Miller, 2019). In comparison, in *Arabidopsis*, only a few histone modifications and histone variants have been implicated in repair and their mechanisms of action are yet to be deciphered. Some of the key histone modifications implicated in DNA repair in these organisms are discussed below.

Histone Phosphorylation: One of the best characterized histone post-translational modifications (PTMs) involved in the DDR is phosphorylation of the histone variant H2A.X. In both plants and mammals, this histone variant is phosphorylated on Ser139 (γH2A.X) by the ATM and ATR kinases in response to DSBs (J. D. Friesner et al., 2005; Rogakou, Pilch, Orr, Ivanova, & Bonner, 1998). γH2A.X marks >1Mb of chromatin around the DSB in mammals and is commonly used as a DSB marker for immunocytology across species, including plants, as they can be observed as foci in response to DNA damage(Natale et al., 2017; H. Su et al., 2017). In mammals, it has been shown that γH2A.X is bound by the BRCT domain phospho-reader protein MDC1 to promote DDR signaling and the recruitment of repair proteins including 53BP1, BRCA1, and the MRN complex (Miller & Jackson, 2012; Stucki et al., 2005; Turinetto & Giachino, 2015).

Moreover, eraser proteins like the PP2A phosphatase, which dephosphorylates γH2A.X at damage sites, have been identified in mammals(Chowdhury et al., 2005). In plants, the role γH2A.X in promoting plant genome stability in response to genotoxic stresses have been established(J. D. Friesner et al., 2005; Waterworth, Footitt, Bray, Finch-Savage, & West, 2016). However, the readers, erasers, and/or the protein complexes interacting with this modification in plants and how these factors help facilitate DNA repair is yet to be deciphered.

In addition to γH2A.X, several other phosphorylation marks have been implicated in DNA repair, and the enzymes, as well as pathways that modulate these marks are emerging in mammals. For instance, phosphorylation of H2BS14 is involved in DSB repair (Fernandez-Capetillo, Allis, & Nussenzweig, 2004) and phosphorylation of H3S10, H3T11, H3S28 are implicated in mitosis, chromatin compaction, and DNA damage pathways(Ozawa, 2008; Sharma, Bhattacharya, Khan, Khade, & Gupta, 2015).

Histone acetylation (ac): Histones are acetylated by the addition of an acetyl group onto lysine residues by the histone acetyl transferase (HAT) family of writer proteins, while these marks are removed by the histone deacetylase (HDAC) family of eraser proteins. Proteins involved in histone acetylation, including the bromo- and YEATS- domain histone acylation readers, have been shown to be important for DSB repair across species(Chiu, Gong, & Miller, 2017; F. Gong et al., 2015; F. Gong, Chiu, & Miller, 2016; D. Zhao, Li, Xiong, Chen, & Li, 2017). In mammals, several histone acetylation marks, including H2A.XK5ac(Ikura et al., 2007), H2A.XK36ac(Jiang et al., 2010), H2BK120ac(Clouaire et al., 2018), H3K9ac(McCord et al., 2009; Tjeertes, Miller, & Jackson, 2009), H3K56ac(Das, Lucia, Hansen, & Tyler, 2009; Tjeertes et al., 2009), etc. have been shown to be involved in altering chromatin structure at DNA damage sites as well as providing recruitment platforms for DNA repair factors. These aforementioned key roles in DNA repair are

in line with the role of histone acetylation in chromatin decompaction and making the chromatin more accessible.

Histone methylation (me): Histone methylation occurs by the addition of one or more methyl group onto lysine residues (mono- (me), di- (me2), or tri- (me3) methylated) or arginine residues (me or me2). Histone methylation levels are maintained *in vivo* by histone methyl transferases and histone demethylases and these marks are important in maintaining either active and silenced transcriptional states(F. Gong & Miller, 2019; Khurana et al., 2014). The demethylation of an active transcription mark, H3K4me3 has been shown to occur at DNA damage sites and is involved in transcriptional repression following DSB formation(F. Gong & Miller, 2019; Hendriks, Treffers, Verlaan-de Vries, Olsen, & Vertegaal, 2015; Mosammaparast et al., 2013). On the other hand, the repressive chromatin mark H3K9me3(Barski et al., 2007) was identified at DSBs, and shown to provide a binding site for Heterochromatin Protein 1 (HP1) and associated proteins, which are important to inactivate transcription(Ayrapetov, Gursoy-Yuzugullu, Xu, Xu, & Price, 2014). In addition, the repressive H3K27me3 mark was shown to accumulate at DSBs(O'Hagan, Mohammad, & Baylin, 2008) and H3K36me3 marks were shown to aid DSB repair via HR(Aymard et al., 2014). This evidence, in addition to the accumulation of several methyl transferases and demethylases at DNA damage sites in mammals, suggests that histone methylation is a dynamic chromatin mark at sites of DNA damage(F. Gong & Miller, 2019; Khurana et al., 2014).

Histone ubiquitylation (ub): Ubiquitylation involves the addition of ubiquitin, a highly conserved 76-amino-acid protein, onto a lysine amino acid residue in a substrate protein, including histones(Pickart, 2001). While K48-linked polyubiquitin chains are associated with proteasomal degradation of the target proteins, K63-linked polyubiquitin chains regulate cellular processes

including the DDR in a degradation-independent manner(Jackson & Durocher, 2013). Histone H2A is ubiquitinylated (on multiple lysines) by PRC1 after DNA damage and is responsible for transcriptional repression around DNA damage sites(Buchwald et al., 2006; Z. Gao et al., 2012; Kakarougkas et al., 2014; H. Wang et al., 2004). H2AK15ub, in coordination with H4K20me2, recruits the histone reader, 53BP1, which is involved in DSB repair and determines the choice between NHEJ and HR repair pathways(Botuyan et al., 2006; Fradet-Turcotte et al., 2013; Huyen et al., 2004; Panier & Boulton, 2014). Finally, BRCA1/BARD1-mediated H2AK127/129ub has been shown to be an important modification to promote DNA end resection, a crucial step deciding repair pathway choice.(Kalb, Mallery, Larkin, Huang, & Hiom, 2014; Uckelmann et al., 2018). This evidence suggest that histone ubiquitination is important for recruitment of key repair factors and thereby influences the repair pathway choice between HR and NHEJ.

In addition to ubiquitin, a related modification, small ubiquitin-related modifier (SUMO), is also added to lysine residues and has been shown to accumulate at DNA damage sites in mammalian cells. Like ubiquitination, SUMO modifications are also important for promoting DSB repair by regulating the recruitment of repair factors such as BRCA1 and 53BP1 to DNA damage sites(Galanty et al., 2009; Jackson & Durocher, 2013; Morris et al., 2009).

Histone variants: Incorporation of histone variant, H2A.Z disrupts nucleosome stability and has been implicated in DNA repair(Horigome et al., 2014; Nishibuchi et al., 2014; Y. Xu et al., 2012) and transcriptional regulation (both activation and repression)(Lashgari, Millau, Jacques, & Gaudreau, 2017; Valdes-Mora et al., 2012; Weber, Ramachandran, & Henikoff, 2014). Moreover, incorporation of H2A.Z into DNA damage sites promotes further chromatin modifications at break sites, which in turn recruits repair factors(Gursoy-Yuzugullu, Ayrapetov, & Price, 2015; Y. Xu et al., 2012). H2A.Z PTMs have been identified after DNA repair.

SUMOylation of H2A.Z (H2A.Zsu) was shown to be required for the recruitment of DSBs to the nuclear periphery in yeast(Kalocsay, Hiller, & Jentsch, 2009). Moreover, H2A.Z.2su in mammalian cells was implicated in DNA repair by facilitating DNA damage dependent exchange of H2A.Z at breaks(Fukuto et al., 2018). H2A.Z is involved in several stress responses, ranging from heat stress to DNA damage. Further investigation into stress specific PTMs on H2A.Z might shed light upon the versatile roles of this variant.

Overall, current evidence suggests that various histone variants and modifications act in a collaborative manner to form a repair conducive environment and play roles in various aspects of signaling and repairing DNA damage. However, these damage-implicated modifications represent a mere fraction of the chromatin modifications identified so far. With the use of advanced NGS and biochemical approaches (Systematic ChIP-seq of multiple histone modifications(Clouaire et al., 2018), PICh- proteomics of isolated chromatin segments(Y. Gong, Handa, Kowalczykowski, & de Lange, 2017) etc.), a more clear picture of the chromatin landscape at the DSBs and how they influence repair is emerging. While the reader, writer and eraser enzymes for some of these modifications (which allows these marks to be reversible and dynamic) have been identified, the complete picture of how chromatin mediates DNA repair across organisms is still not well understood. Characterization of *Arabidopsis* proteins containing reader, writer and eraser domains that interact with the DNA damage-associated chromatin modifications will help understand the role of chromatin in plant DDR.

Chromatin modifiers in DNA Repair: Insights from *Arabidopsis*

Compared to mammals and yeast, knowledge about the roles of chromatin modifications in DNA repair remains limited in *Arabidopsis*. Although only a few chromatin modifications have

been connected to DNA repair in plants, homologs of some chromatin modifiers such as histone readers, writers, and erasers, along with several chromatin remodeling and histone chaperone complexes, and even a few different types of non-coding RNA, have been implicated in DNA repair as summarized below.

Histone readers, writers and erasers: Chromatin modifications influence DNA processes, including DNA repair, via interactions with reader, writer, and eraser proteins. However, a systematic study of histone readers in *Arabidopsis*, has been only recently performed(S. Zhao, Zhang, Yang, Zhu, & Li, 2018). Of these, only two histone readers are currently implicated in DNA repair and associated processes. MUTS HOMOLOG 6 (MSH6), an H3K4me3-reader and component of the post-replicative DNA mismatch repair system (MMR)(S. Zhao et al., 2018), has been shown to play roles in somatic recombination(Gonzalez & Spampinato, 2020) and cadmium-induced checkpoint arrest(Cao et al., 2018). Moreover, MSH2, which forms a heterodimer with MSH6, has been shown to shape meiotic crossover landscape(Blackwell et al., 2020). Various histone writers, such histone methyltransferases and acetyltransferases have also been implicated in *Arabidopsis* DNA repair. The H3K27 methyltransferase, CURLY LEAF (CLF), is associated with the epigenetic regulation of somatic HR repair(N. Chen, Zhou, Wang, Dong, & Yu, 2014). The H3K27 mono-methyltransferases, ARABIDOPSIS TRITHORAX-RELATED5 (ATXR5) and ATXR6, have been shown to be required for chromatin structure and gene silencing(Feng et al., 2017; Jacob et al., 2014). The histone acetyltransferases HAIRY MERISTEM 1 (HAM1) and HISTONE ACETYLTRANSFERASE OF THE GNAT FAMILY 3 (HAG3), were shown to participate in UV-B-induced DDR signaling and DNA repair by negatively regulating the expression of DNA repair enzymes(Campi, D'Andrea, Emiliani, & Casati, 2012; Fina & Casati, 2015). Histone Acetyltransferases HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY

1 (HAC1) and HISTONE ACETYLTRANSFERASE OF THE TAFII250 FAMILY 1 (HAF1), have crucial roles in UV-B signaling (Fina, Masotti, Rius, Crevacuore, & Casati, 2017). Finally, some histone eraser proteins, such as histone deacetylases have also been implicated in DNA repair-associated processes in *Arabidopsis.* HISTONE DEACETYLASE 2 (HDA2), HDA6, and HDA19 were shown to modulate levels of H3K9K14ac or H3K9ac and are important for gene regulation in abiotic stress responses, pathways that crosstalk with DDR pathways(Luo, Cheng, Xu, Yang, & Wu, 2017).

Histone PTMs are extensively conserved across organisms and therefore, by analogy to other organisms, there are likely many other chromatin readers, writers and erasers with important roles in DNA repair that once identified, would begin filling a major gap in our understanding of DSB repair in plants.

Chromatin Remodelers: To create a repair conducive environment and provide repair complexes access to any genomic location that could be plausibly damaged, chromatin remodeling factors can slide, evict, or even exchange nucleosomes throughout the genome. The *Arabidopsis* SWR1 complex is involved in the substitution of H2A for H2A.Z in nucleosomes(March-Diaz, Garcia-Dominguez, Florencio, & Reyes, 2007; Noh & Amasino, 2003) and its subunits, PIE1, ARP6, and SWC6 are have been implicated in somatic HR repair and meiosis(Rosa, Von Harder, Cigliano, Schlogelhofer, & Mittelsten Scheid, 2013). Moreover, SWR1 subunits and H2A.Z were shown to have non-redundant functions in plant immunity and gene regulation(Berriri, Gangappa, & Kumar, 2016), processes that have close ties to the DDR in plants. INO80 was shown to play roles in both transcription and HR-like repair of DNA damage(Fritsch et al., 2004) and was shown to be crucial for genome stability and plant development(C. Zhang et al., 2015). The SWI2/SNF2 family chromatin remodeler RAD54, is

important for DSB repair via HR and interacts with another DNA repair protein, RAD51(Osakabe et al., 2006; Shaked, Avivi-Ragolsky, & Levy, 2006). Another SWI2/SNF2 family chromatin remodeler, DDM1 contributes to the methylation and stable silencing of transposable elements by allowing DNA methyltransferases to access H1-containing heterochromatin(Zemach et al., 2013). *ddm1* mutants display reduced chromatin accessibility and HR-like repair defects after DSB damage(Choi et al., 2019). The SMC complex sub-components, MMS21 and NSE4 were shown to function in HR-like repair(Diaz & Pecinka, 2018; P. Xu et al., 2013; Yuan et al., 2014) and to be important for tolerance of zebularine-induced DNA damage (Diaz et al., 2019), respectively. This evidence demonstrates that many different remodeling complexes are required for repair, but their precise roles remain to be determined.

Histone chaperones: Histone chaperones prevent undesirable non-specific interactions between DNA and highly basic histone proteins before they assemble to form the nucleosome. Various H2A-H2B and H3-H4 histone chaperones have been implicated in DNA repair in *Arabidopsis*. The H2A-H2B chaperones NAP1, NAP1-RELATED PROTEIN (NRP1), and NRP2 promote nucleosome disassembly and reassembly during HR(J. Gao et al., 2012). The H2A-H2B FACT (facilitates chromatin transcription) chaperone functions as a key protein in regulating homologous recombination via chromatin remodeling(Oliveira et al., 2014) and is also required for DNA demethylation at heterochromatin via DEMETER (DME) during reproduction(Frost et al., 2018). CHROMATIN ASSEMBLY FACTOR-1 (CAF-1) facilitates the incorporation of histones H3 and H4 onto newly synthesized DNA and is required for expression of key DDR genes, mitotic chromosome integrity(Varas, Santos, & Pradillo, 2017), and heterochromatin formation(Schonrock, Exner, Probst, Gruissem, & Hennig, 2006). The ANTI-SILENCING FUNCTION PROTEIN 1 (ASF1A and ASF1B) chaperones bind histone H3 and are required for

chromatin replication, maintenance of genome integrity, and cell proliferation in a redundant-manner(Zhu et al., 2011). Finally, the H3–H4 chaperone, HISTONE CELL CYCLE REGULATION DEFECTIVE HOMOLOG A (HIRA) that deposits histone H3.3 into chromatin, was shown to influence transcriptional dynamics and impact environmental stress response(Nie, Wang, Li, Holec, & Berger, 2014). In line with the role of histone chaperones in nucleosome assembly, in the context of DNA repair, they are expected to provide access to repair proteins and rebuild chromatin after repair. In agreement with this expectation, mutants of subunits in chaperone complexes show roles in repair, but their precise roles mediated by chromatin are yet to be determined.

DNA methylation: DNA methylation is a highly conserved chromatin modification that is associated with gene silencing and plays critical roles in development and maintaining genome stability through the suppression of transposable elements. The DNA demethylase, REPRESSOR OF SILENCING1 (ROS1) was shown to be required for genotoxic stress tolerance as a repressor of transcriptional gene silencing(Z. Gong et al., 2002). Moreover, a DNA repair factor belonging to the UV-damaged DNA-binding protein complex, DNA DAMAGE BINDING PROTEIN2 (DDB2), was shown to represses enzymatic activity of demethylase ROS1(Cordoba-Canero, Cognat, Ariza, Roldan Arjona, & Molinier, 2017), and also to transcriptionally regulate the expression of the *ROS1* and *DEMETER-LIKE PROTEIN 3 (DML3)* demethylases. In addition, DDB2 also influences *de novo* DNA methylation by forming functional DDB2-AGO4-small interfering RNA (siRNA) complexes(Cordoba-Canero et al., 2017; Schalk et al., 2017). Together, these findings demonstrate roles for DNA methylation in repair via transcriptional repression of key genes. However, the chromatin-mediated role of DNA methylation in repair is unknown. Given the importance of maintaining an accessible chromatin environment at the break site to

promote DNA repair, studying how the repressive methylation marks are removed around the break site is a promising area for further research.

Non-coding RNA: Recent progress in high-throughput sequencing using next-generation sequencing (NGS) technologies has enabled the identification of various classes of non-coding RNAs (ncRNAs): microRNAs (miRNAs), small RNAs (sRNAs), and long non-coding RNAs (lncRNAs). Various ncRNA have emerged as essential regulators of plants abiotic stress responses, including DNA damage repair, and our understanding of how these RNAs are involved in the DDR are summarized below.

Several long non-coding RNAs (lncRNA), some of which are induced by DNA damage, have shown to be involved in the DDR and in maintaining genomic integrity in mammals(Statello, Guo, Chen, & Huarte, 2021). They participate in the DDR in a diverse manner, including guiding proteins to specific sites in the genome, holding broken ends together, mediating DNA-protein interactions, signaling DNA damage, acting as a scaffold, recruiting repair proteins, and initiating chromatin reconstruction(Dianatpour & Ghafouri-Fard, 2017; Durut & Mittelsten Scheid, 2019; Guiducci & Stojic, 2021; Statello et al., 2021). While the mechanistic roles of lncRNA in DDR remain relatively unexplored in plants, several lncRNAs that are differentially expressed after DSB stress have been identified(Z. Wang et al., 2016). Of these lncRNAs, *TER2* was identified as a novel DNA damage-induced noncoding RNA that works in concert with the *TER1* to promote genome integrity at telomeres(Cifuentes-Rojas et al., 2012). Further biochemical characterization of the differentially expressed lncRNA identified in plants will perhaps provide insights into how lncRNA might be regulating the DDR in plants.

miRNAs are important regulators of gene expression in plant stress responses as well as plant growth, development, and maintenance of genome integrity(Khraiwesh, Zhu, & Zhu, 2012;

Waititu et al., 2020). While various miRNA have been shown to be responsive to UV-B (miR156/159/160/166/390/393/398) and γ-IR (miR840/850), and to participate in the regulatory network of plant stress responses, the mechanisms through which these miRNA influence DDR are yet to be identified(J. H. Kim, Go, Kim, & Chung, 2016; Sunkar, Kapoor, & Zhu, 2006; X. Zhou, Wang, & Zhang, 2007).

In *Arabidopsis*, different sRNA classes have been shown to have roles in DNA repair. The 24-nt small interfering RNAs (siRNAs) are responsible for *de novo* methylation through a process called RNA-directed DNA methylation (RdDM) in plants. Upon DNA stress, the RdDM machinery has been shown to produce 21-nt sRNAs(Schalk et al., 2017; Wei et al., 2012). After UV damage, these 21-nt sRNAs are required for the recognition and repair of DNA photoproducts by forming a chromatin-bound complex with DDB2 and AGO1(Schalk et al., 2017). Upon DSB induction in transgenic lines, 21-nt sRNAs (named DSB-induced small RNAs or diRNAs) are also produced, recruited at DSB sites via interaction with AGO2 to repair lesions, and are required for DSB repair or DDR activation(d'Adda di Fagagna, 2014; Kleinboelting, Huep, Kloetgen, Viehoever, & Weisshaar, 2012). However, at endogenous loci, diRNAs are not necessary for DSB repair(Miki et al., 2017). Moreover, ARGONAUTE proteins, which are responsible for targeting the sRNA, have roles in DNA repair and meiosis(Bajczyk et al., 2019; Oliver, Santos, & Pradillo, 2014). These evidence, showing complex interconnections between sRNA machinery and DNA repair pathways, widens the scope of sRNAs in genome maintenance.

THESIS OBJECTIVES

The study of DNA damage repair pathways is crucial, as it has been implicated in diverse processes, ranging from ageing and cancer in animals, to decreased crop yields in plants. As outlined in the previous sections, even though some key regulators and factors participating in plant DNA repair have been identified, the pathways and mechanisms through which they regulate the DDR processes via chromatin is yet to be elucidated. Two critical, but as yet poorly understood aspects of the DDR in the plant model, *Arabidopsis thaliana*, are (**1**) the kinetics and regulatory networks controlling the expression of genes involved in the key biological processes during the DDR and (**2**) the roles of specific chromatin modifications, and the pathways they are implicated in, to regulate the process of DNA repair. The work in this thesis addresses the aforementioned questions. **Chapter two** presents the development of a temporal model for the *Arabidopsis* transcriptome after DNA damage. This work, using genetic, genomic, and biochemical approaches, mapped out the dynamics and regulation of key DDR processes over a 24-hour time course. **Chapter three** presents the setup of a reverse genetic screen to identify candidate chromatin-associated factors in repair, and the characterization of the role of a candidate chromatin reader, YAF9B and its homolog YAF9A in DSB repair via chromatin-mediated pathways. Together, these chapters broaden and deepen the understanding of the role of the DDR transcriptional networks and chromatin in orchestrating DNA repair by revealing the dynamics of transcriptional regulation during the DNA damage response and roles for select histone reader proteins in DNA repair.

REFERENCES

Adachi, S., Minamisawa, K., Okushima, Y., Inagaki, S., Yoshiyama, K., Kondou, Y., . . . Umeda, M. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. *Proc Natl Acad Sci U S A, 108*(24), 10004-10009. doi:10.1073/pnas.1103584108

Aguilera, A., & Garcia-Muse, T. (2013). Causes of genome instability. *Annu Rev Genet, 47*, 1-32. doi:10.1146/annurev-genet-111212-133232

Amiard, S., Gallego, M. E., & White, C. I. (2013). Signaling of double strand breaks and deprotected telomeres in Arabidopsis. *Front Plant Sci, 4*, 405. doi:10.3389/fpls.2013.00405

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., . . . Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol, 21*(4), 366-374. doi:10.1038/nsmb.2796

Ayrapetov, M. K., Gursoy-Yuzugullu, O., Xu, C., Xu, Y., & Price, B. D. (2014). DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proc Natl Acad Sci U S A, 111*(25), 9169-9174. doi:10.1073/pnas.1403565111

Bagherieh-Najjar, M. B., de Vries, O. M., Hille, J., & Dijkwel, P. P. (2005). Arabidopsis RecQI4A suppresses homologous recombination and modulates DNA damage responses. *Plant J, 43*(6), 789-798. doi:10.1111/j.1365-313X.2005.02501.x

Bajczyk, M., Bhat, S. S., Szewc, L., Szweykowska-Kulinska, Z., Jarmolowski, A., & Dolata, J. (2019). Novel Nuclear Functions of Arabidopsis ARGONAUTE1: Beyond RNA Interference. *Plant Physiol, 179*(3), 1030-1039. doi:10.1104/pp.18.01351

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res, 21*(3), 381-395. doi:10.1038/cr.2011.22

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837. doi:10.1016/j.cell.2007.05.009

Bartek, J., Falck, J., & Lukas, J. (2001). CHK2 kinase--a busy messenger. *Nat Rev Mol Cell Biol, 2*(12), 877-886. doi:10.1038/35103059

Beck, C. B. (2010). *An introduction to plant structure and development : plant anatomy for the twenty-first century* (2nd ed.). Cambridge ; New York: Cambridge University Press.

Berriri, S., Gangappa, S. N., & Kumar, S. V. (2016). SWR1 Chromatin-Remodeling Complex Subunits and H2A.Z Have Non-overlapping Functions in Immunity and Gene Regulation in Arabidopsis. *Mol Plant, 9*(7), 1051-1065. doi:10.1016/j.molp.2016.04.003

Biedermann, S., Harashima, H., Chen, P., Heese, M., Bouyer, D., Sofroni, K., & Schnittger, A. (2017). The retinoblastoma homolog RBR1 mediates localization of the repair protein RAD51 to DNA lesions in Arabidopsis. *EMBO J, 36*(9), 1279-1297. doi:10.15252/embj.201694571

Biever, J. J., Brinkman, D., & Gardner, G. (2014). UV-B inhibition of hypocotyl growth in etiolated Arabidopsis thaliana seedlings is a consequence of cell cycle arrest initiated by photodimer accumulation. *J Exp Bot, 65*(11), 2949-2961. doi:10.1093/jxb/eru035

Blackwell, A. R., Dluzewska, J., Szymanska-Lejman, M., Desjardins, S., Tock, A. J., Kbiri, N., . . . Henderson, I. R. (2020). MSH2 shapes the meiotic crossover landscape in relation to interhomolog polymorphism in Arabidopsis. *EMBO J, 39*(21), e104858. doi:10.15252/embj.2020104858

Bleuyard, J. Y., Gallego, M. E., & White, C. I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair (Amst), 5*(1), 1-12. doi:10.1016/j.dnarep.2005.08.017

Block-Schmidt, A. S., Dukowic-Schulze, S., Wanieck, K., Reidt, W., & Puchta, H. (2011). BRCC36A is epistatic to BRCA1 in DNA crosslink repair and homologous recombination in Arabidopsis thaliana. *Nucleic Acids Res, 39*(1), 146-154. doi:10.1093/nar/gkq722

Botuyan, M. V., Lee, J., Ward, I. M., Kim, J. E., Thompson, J. R., Chen, J., & Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell, 127*(7), 1361-1373. doi:10.1016/j.cell.2006.10.043

Breuer, C., Morohashi, K., Kawamura, A., Takahashi, N., Ishida, T., Umeda, M., . . . Sugimoto, K. (2012). Transcriptional repression of the APC/C activator CCS52A1 promotes active termination of cell growth. *EMBO J, 31*(24), 4488-4501. doi:10.1038/emboj.2012.294

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., & Sixma, T. K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J, 25*(11), 2465-2474. doi:10.1038/sj.emboj.7601144

Buschbeck, M., & Hake, S. B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat Rev Mol Cell Biol, 18*(5), 299-314. doi:10.1038/nrm.2016.166

Caldecott, K. W. (2008). Single-strand break repair and genetic disease. *Nat Rev Genet, 9*(8), 619-631. doi:10.1038/nrg2380

Campi, M., D'Andrea, L., Emiliani, J., & Casati, P. (2012). Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. *Plant Physiol, 158*(2), 981-995. doi:10.1104/pp.111.191452

Cao, X., Wang, H., Zhuang, D., Zhu, H., Du, Y., Cheng, Z., . . . Liu, W. (2018). Roles of MSH2 and MSH6 in cadmium-induced G2/M checkpoint arrest in Arabidopsis roots. *Chemosphere, 201*, 586-594. doi:10.1016/j.chemosphere.2018.03.017

Charbonnel, C., Allain, E., Gallego, M. E., & White, C. I. (2011). Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. *DNA Repair (Amst), 10*(6), 611-619. doi:10.1016/j.dnarep.2011.04.002

Charbonnel, C., Gallego, M. E., & White, C. I. (2010). Xrcc1-dependent and Ku-dependent DNA double-strand break repair kinetics in Arabidopsis plants. *Plant J, 64*(2), 280-290. doi:10.1111/j.1365-313X.2010.04331.x

Chen, N., Zhou, W. B., Wang, Y. X., Dong, A. W., & Yu, Y. (2014). Polycomb-group histone methyltransferase CLF is required for proper somatic recombination in Arabidopsis. *J Integr Plant Biol, 56*(6), 550-558. doi:10.1111/jipb.12157

Chen, P., & Umeda, M. (2015). DNA double-strand breaks induce the expression of flavin-containing monooxygenase and reduce root meristem size in Arabidopsis thaliana. *Genes Cells, 20*(8), 636-646. doi:10.1111/gtc.12255

Chen, Y., & Sanchez, Y. (2004). Chk1 in the DNA damage response: conserved roles from yeasts to mammals. *DNA Repair (Amst), 3*(8-9), 1025-1032. doi:10.1016/j.dnarep.2004.03.003

Chiu, L. Y., Gong, F., & Miller, K. M. (2017). Bromodomain proteins: repairing DNA damage within chromatin. *Philos Trans R Soc Lond B Biol Sci, 372*(1731). doi:10.1098/rstb.2016.0286

Choi, S. H., Ryu, T. H., Kim, J. I., Lee, S., Lee, S. S., & Kim, J. H. (2019). Mutation in DDM1 inhibits the homology directed repair of double strand breaks. *Plos One, 14*(2), e0211878. doi:10.1371/journal.pone.0211878

Chowdhury, D., Keogh, M. C., Ishii, H., Peterson, C. L., Buratowski, S., & Lieberman, J. (2005). gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. *Mol Cell, 20*(5), 801-809. doi:10.1016/j.molcel.2005.10.003

Ciccia, A., & Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. *Mol Cell, 40*(2), 179-204. doi:10.1016/j.molcel.2010.09.019

Cifuentes-Rojas, C., Nelson, A. D., Boltz, K. A., Kannan, K., She, X., & Shippen, D. E. (2012). An alternative telomerase RNA in Arabidopsis modulates enzyme activity in response to DNA damage. *Genes Dev, 26*(22), 2512-2523. doi:10.1101/gad.202960.112

Clouaire, T., Rocher, V., Lashgari, A., Arnould, C., Aguirrebengoa, M., Biernacka, A., . . . Legube, G. (2018). Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell, 72*(2), 250-262 e256. doi:10.1016/j.molcel.2018.08.020

Cools, T., & De Veylder, L. (2009). DNA stress checkpoint control and plant development. *Curr Opin Plant Biol, 12*(1), 23-28. doi:10.1016/j.pbi.2008.09.012

Cordoba-Canero, D., Cognat, V., Ariza, R. R., Roldan Arjona, T., & Molinier, J. (2017). Dual control of ROS1-mediated active DNA demethylation by DNA damage-binding protein 2 (DDB2). *Plant J, 92*(6), 1170-1181. doi:10.1111/tpj.13753

Culligan, K. M., Robertson, C. E., Foreman, J., Doerner, P., & Britt, A. B. (2006). ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J, 48*(6), 947-961. doi:10.1111/j.1365-313X.2006.02931.x

d'Adda di Fagagna, F. (2014). A direct role for small non-coding RNAs in DNA damage response. *Trends Cell Biol, 24*(3), 171-178. doi:10.1016/j.tcb.2013.09.008

Das, C., Lucia, M. S., Hansen, K. C., & Tyler, J. K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature, 459*(7243), 113-117. doi:10.1038/nature07861

De Schutter, K., Joubes, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., . . . De Veylder, L. (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. *Plant Cell, 19*(1), 211-225. doi:10.1105/tpc.106.045047

Dianatpour, A., & Ghafouri-Fard, S. (2017). The Role of Long Non Coding RNAs in the Repair of DNA Double Strand Breaks. *Int J Mol Cell Med, 6*(1), 1-12.

Diaz, M., & Pecinka, A. (2018). Scaffolding for Repair: Understanding Molecular Functions of the SMC5/6 Complex. *Genes (Basel), 9*(1). doi:10.3390/genes9010036

Diaz, M., Pecinkova, P., Nowicka, A., Baroux, C., Sakamoto, T., Gandha, P. Y., . . . Pecinka, A. (2019). The SMC5/6 Complex Subunit NSE4A Is Involved in DNA Damage Repair and Seed Development. *Plant Cell, 31*(7), 1579-1597. doi:10.1105/tpc.18.00043

Dijkwel, P. P., & Lai, A. G. (2019). Hypothesis: Plant stem cells hold the key to extreme longevity. *Translational Medicine of Aging, 3*, 14-16. doi:10.1016/j.tma.2018.12.002

Ding, L. H., Park, S., Peyton, M., Girard, L., Xie, Y., Minna, J. D., & Story, M. D. (2013). Distinct transcriptome profiles identified in normal human bronchial epithelial cells after exposure to gamma-rays and different elemental particles of high Z and energy. *BMC Genomics, 14*, 372. doi:10.1186/1471-2164-14-372

Doutriaux, M. P., Couteau, F., Bergounioux, C., & White, C. (1998). Isolation and characterisation of the RAD51 and DMC1 homologs from Arabidopsis thaliana. *Mol Gen Genet, 257*(3), 283-291. doi:10.1007/s004380050649

Durrant, W. E., Wang, S., & Dong, X. (2007). Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. *Proc Natl Acad Sci U S A, 104*(10), 4223-4227. doi:10.1073/pnas.0609357104

Durut, N., & Mittelsten Scheid, O. (2019). The Role of Noncoding RNAs in Double-Strand Break Repair. *Front Plant Sci, 10*, 1155. doi:10.3389/fpls.2019.01155

Einset, J., & Collins, A. R. (2015). DNA repair after X-irradiation: lessons from plants. *Mutagenesis, 30*(1), 45-50. doi:10.1093/mutage/geu054

Feng, W., Hale, C. J., Over, R. S., Cokus, S. J., Jacobsen, S. E., & Michaels, S. D. (2017). Large-scale heterochromatin remodeling linked to overreplication-associated DNA damage. *Proc Natl Acad Sci U S A, 114*(2), 406-411. doi:10.1073/pnas.1619774114

Fernandez-Capetillo, O., Allis, C. D., & Nussenzweig, A. (2004). Phosphorylation of histone H2B at DNA double-strand breaks. *J Exp Med, 199*(12), 1671-1677. doi:10.1084/jem.20032247

Fina, J. P., & Casati, P. (2015). HAG3, a Histone Acetyltransferase, Affects UV-B Responses by Negatively Regulating the Expression of DNA Repair Enzymes and Sunscreen Content in Arabidopsis thaliana. *Plant Cell Physiol, 56*(7), 1388-1400. doi:10.1093/pcp/pcv054

Fina, J. P., Masotti, F., Rius, S. P., Crevacuore, F., & Casati, P. (2017). HAC1 and HAF1 Histone Acetyltransferases Have Different Roles in UV-B Responses in Arabidopsis. *Front Plant Sci, 8*, 1179. doi:10.3389/fpls.2017.01179

Fox, D. T., & Duronio, R. J. (2013). Endoreplication and polyploidy: insights into development and disease. *Development, 140*(1), 3-12. doi:10.1242/dev.080531

Fradet-Turcotte, A., Canny, M. D., Escribano-Diaz, C., Orthwein, A., Leung, C. C., Huang, H., . . . Durocher, D. (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature, 499*(7456), 50-54. doi:10.1038/nature12318

Friedberg, E. C., Walker, G. C., Siede, W., & Wood, R. D. (2005). *DNA repair and mutagenesis*: American Society for Microbiology Press.

Friesner, J., & Britt, A. B. (2003). Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *Plant J, 34*(4), 427-440. doi:10.1046/j.1365-313x.2003.01738.x

Friesner, J. D., Liu, B., Culligan, K., & Britt, A. B. (2005). Ionizing radiation-dependent gamma-H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. *Mol Biol Cell, 16*(5), 2566-2576. doi:10.1091/mbc.e04-10-0890

Fritsch, O., Benvenuto, G., Bowler, C., Molinier, J., & Hohn, B. (2004). The INO80 protein controls homologous recombination in Arabidopsis thaliana. *Mol Cell, 16*(3), 479-485. doi:10.1016/j.molcel.2004.09.034

Frost, J. M., Kim, M. Y., Park, G. T., Hsieh, P. H., Nakamura, M., Lin, S. J. H., . . . Fischer, R. L. (2018). FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc Natl Acad Sci U S A, 115*(20), E4720-E4729. doi:10.1073/pnas.1713333115

Fukuto, A., Ikura, M., Ikura, T., Sun, J., Horikoshi, Y., Shima, H., . . . Tashiro, S. (2018). SUMO modification system facilitates the exchange of histone variant H2A.Z-2 at DNA damage sites. *Nucleus, 9*(1), 87-94. doi:10.1080/19491034.2017.1395543

Fulcher, N., & Sablowski, R. (2009). Hypersensitivity to DNA damage in plant stem cell niches. *Proc Natl Acad Sci U S A, 106*(49), 20984-20988. doi:10.1073/pnas.0909218106

Furukawa, T., Curtis, M. J., Tominey, C. M., Duong, Y. H., Wilcox, B. W., Aggoune, D., . . . Britt, A. B. (2010). A shared DNA-damage-response pathway for induction of stem-cell death by UVB and by gamma irradiation. *DNA Repair (Amst), 9*(9), 940-948. doi:10.1016/j.dnarep.2010.06.006

Galanty, Y., Belotserkovskaya, R., Coates, J., Polo, S., Miller, K. M., & Jackson, S. P. (2009). Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature, 462*(7275), 935-939. doi:10.1038/nature08657

Gao, J., Zhu, Y., Zhou, W., Molinier, J., Dong, A., & Shen, W. H. (2012). NAP1 family histone chaperones are required for somatic homologous recombination in Arabidopsis. *Plant Cell, 24*(4), 1437-1447. doi:10.1105/tpc.112.096792

Gao, M., Wei, W., Li, M. M., Wu, Y. S., Ba, Z., Jin, K. X., . . . Rendtlew Danielsen, J. M. (2014). Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Res, 24*(5), 532-541. doi:10.1038/cr.2014.36

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., . . . Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell, 45*(3), 344-356. doi:10.1016/j.molcel.2012.01.002

Gong, F., Chiu, L. Y., Cox, B., Aymard, F., Clouaire, T., Leung, J. W., . . . Miller, K. M. (2015). Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes Dev, 29*(2), 197-211. doi:10.1101/gad.252189.114

Gong, F., Chiu, L. Y., & Miller, K. M. (2016). Acetylation Reader Proteins: Linking Acetylation Signaling to Genome Maintenance and Cancer. *PLoS Genet, 12*(9), e1006272. doi:10.1371/journal.pgen.1006272

Gong, F., & Miller, K. M. (2019). Histone methylation and the DNA damage response. *Mutat Res Rev Mutat Res, 780*, 37-47. doi:10.1016/j.mrrev.2017.09.003

Gong, Y., Handa, N., Kowalczykowski, S. C., & de Lange, T. (2017). PHF11 promotes DSB resection, ATR signaling, and HR. *Genes Dev, 31*(1), 46-58. doi:10.1101/gad.291807.116

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T., David, L., & Zhu, J. K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell, 111*(6), 803-814. doi:10.1016/s0092-8674(02)01133-9

Gonzalez, V., & Spampinato, C. P. (2020). The mismatch repair protein MSH6 regulates somatic recombination in Arabidopsis thaliana. *DNA Repair (Amst), 87*, 102789. doi:10.1016/j.dnarep.2020.102789

Gu, Z., Pan, W., Chen, W., Lian, Q., Wu, Q., Lv, Z., . . . Ge, X. (2019). New perspectives on the plant PARP family: Arabidopsis PARP3 is inactive, and PARP1 exhibits predominant poly (ADP-ribose) polymerase activity in response to DNA damage. *BMC Plant Biol, 19*(1), 364. doi:10.1186/s12870-019-1958-9

Guiducci, G., & Stojic, L. (2021). Long Noncoding RNAs at the Crossroads of Cell Cycle and Genome Integrity. *Trends Genet, 37*(6), 528-546. doi:10.1016/j.tig.2021.01.006

Gursoy-Yuzugullu, O., Ayrapetov, M. K., & Price, B. D. (2015). Histone chaperone Anp32e removes H2A.Z from DNA double-strand breaks and promotes nucleosome reorganization and DNA repair. *Proc Natl Acad Sci U S A, 112*(24), 7507-7512. doi:10.1073/pnas.1504868112

Harper, J. W., & Elledge, S. J. (2007). The DNA damage response: ten years after. *Mol Cell, 28*(5), 739-745. doi:10.1016/j.molcel.2007.11.015

Heitzeberg, F., Chen, I. P., Hartung, F., Orel, N., Angelis, K. J., & Puchta, H. (2004). The Rad17 homologue of Arabidopsis is involved in the regulation of DNA damage repair and homologous recombination. *Plant J, 38*(6), 954-968. doi:10.1111/j.1365-313X.2004.02097.x

Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet, 15*(9), 585-598. doi:10.1038/nrg3729

Hendriks, I. A., Treffers, L. W., Verlaan-de Vries, M., Olsen, J. V., & Vertegaal, A. C. O. (2015). SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. *Cell Rep, 10*(10), 1778-1791. doi:10.1016/j.celrep.2015.02.033

Heyman, J., Polyn, S., Eekhout, T., & De Veylder, L. (2017). Tissue-Specific Control of the Endocycle by the Anaphase Promoting Complex/Cyclosome Inhibitors UVI4 and DEL1. *Plant Physiol, 175*(1), 303-313. doi:10.1104/pp.17.00785

Hoeijmakers, J. H. (2009). DNA damage, aging, and cancer. *N Engl J Med, 361*(15), 1475-1485. doi:10.1056/NEJMra0804615

Horigome, C., Oma, Y., Konishi, T., Schmid, R., Marcomini, I., Hauer, M. H., . . . Gasser, S. M. (2014). SWR1 and INO80 chromatin remodelers contribute to DNA double-strand break perinuclear anchorage site choice. *Mol Cell, 55*(4), 626-639. doi:10.1016/j.molcel.2014.06.027

Horvath, B. M., Kourova, H., Nagy, S., Nemeth, E., Magyar, Z., Papdi, C., . . . Scheres, B. (2017). Arabidopsis RETINOBLASTOMA RELATED directly regulates DNA damage responses through functions beyond cell cycle control. *EMBO J, 36*(9), 1261-1278. doi:10.15252/embj.201694561

Hu, Z., Cools, T., & De Veylder, L. (2016). Mechanisms Used by Plants to Cope with DNA Damage. *Annu Rev Plant Biol, 67*, 439-462. doi:10.1146/annurev-arplant-043015-111902

Hu, Z., Cools, T., Kalhorzadeh, P., Heyman, J., & De Veylder, L. (2015). Deficiency of the Arabidopsis helicase RTEL1 triggers a SOG1-dependent replication checkpoint in response to DNA cross-links. *Plant Cell, 27*(1), 149-161. doi:10.1105/tpc.114.134312

Huyen, Y., Zgheib, O., Ditullio, R. A., Jr., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., . . . Halazonetis, T. D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature, 432*(7015), 406-411. doi:10.1038/nature03114

Ikura, T., Tashiro, S., Kakino, A., Shima, H., Jacob, N., Amunugama, R., . . . Kamiya, K. (2007). DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. *Mol Cell Biol, 27*(20), 7028-7040. doi:10.1128/MCB.00579-07

Jackson, S. P., & Durocher, D. (2013). Regulation of DNA damage responses by ubiquitin and SUMO. *Mol Cell, 49*(5), 795-807. doi:10.1016/j.molcel.2013.01.017

Jacob, Y., Bergamin, E., Donoghue, M. T., Mongeon, V., LeBlanc, C., Voigt, P., . . . Martienssen, R. A. (2014). Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science, 343*(6176), 1249-1253. doi:10.1126/science.1248357

Jia, Q., den Dulk-Ras, A., Shen, H., Hooykaas, P. J., & de Pater, S. (2013). Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in Arabidopsis thaliana. *Plant Mol Biol, 82*(4-5), 339-351. doi:10.1007/s11103-013-0065-9

Jiang, Y., Wang, X., Bao, S., Guo, R., Johnson, D. G., Shen, X., & Li, L. (2010). INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proc Natl Acad Sci U S A, 107*(40), 17274-17279. doi:10.1073/pnas.1008388107

Johnson, R. A., Conklin, P. A., Tjahjadi, M., Missirian, V., Toal, T., Brady, S. M., & Britt, A. B. (2018). SUPPRESSOR OF GAMMA RESPONSE1 Links DNA Damage Response to Organ Regeneration. *Plant Physiol, 176*(2), 1665-1675. doi:10.1104/pp.17.01274

Jurgens, G. (2003). Growing up green: cellular basis of plant development. *Mech Dev, 120*(11), 1395-1406. doi:10.1016/j.mod.2003.03.001

Kakarougkas, A., Ismail, A., Chambers, A. L., Riballo, E., Herbert, A. D., Kunzel, J., . . . Downs, J. A. (2014). Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Mol Cell, 55*(5), 723-732. doi:10.1016/j.molcel.2014.06.028

Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T., & Hiom, K. (2014). BRCA1 is a histone-H2A-specific ubiquitin ligase. *Cell Rep, 8*(4), 999-1005. doi:10.1016/j.celrep.2014.07.025

Kalocsay, M., Hiller, N. J., & Jentsch, S. (2009). Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol Cell, 33*(3), 335-343. doi:10.1016/j.molcel.2009.01.016

Khraiwesh, B., Zhu, J. K., & Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta, 1819*(2), 137-148. doi:10.1016/j.bbagrm.2011.05.001

Khurana, S., Kruhlak, M. J., Kim, J., Tran, A. D., Liu, J., Nyswaner, K., . . . Oberdoerffer, P. (2014). A macrohistone variant links dynamic chromatin compaction to BRCA1-dependent genome maintenance. *Cell Rep, 8*(4), 1049-1062. doi:10.1016/j.celrep.2014.07.024

Kim, J. H., Go, Y. S., Kim, J. K., & Chung, B. Y. (2016). Characterization of microRNAs and their target genes associated with transcriptomic changes in gamma-irradiated Arabidopsis. *Genet Mol Res, 15*(3). doi:10.4238/gmr.15038386

Kim, J. J., Lee, S. Y., & Miller, K. M. (2019). Preserving genome integrity and function: the DNA damage response and histone modifications. *Crit Rev Biochem Mol Biol, 54*(3), 208-241. doi:10.1080/10409238.2019.1620676

Kimura, S., & Sakaguchi, K. (2006). DNA repair in plants. *Chem Rev, 106*(2), 753-766. doi:10.1021/cr040482n

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., & Weisshaar, B. (2012). GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res, 40*(Database issue), D1211-1215. doi:10.1093/nar/gkr1047

Knoll, A., Fauser, F., & Puchta, H. (2014). DNA recombination in somatic plant cells: mechanisms and evolutionary consequences. *Chromosome Res, 22*(2), 191-201. doi:10.1007/s10577-014-9415-y

Lashgari, A., Millau, J. F., Jacques, P. E., & Gaudreau, L. (2017). Global inhibition of transcription causes an increase in histone H2A.Z incorporation within gene bodies. *Nucleic Acids Res, 45*(22), 12715-12722. doi:10.1093/nar/gkx879

Law, J. A., & Jacobsen, S. E. (2009). Molecular biology. Dynamic DNA methylation. *Science, 323*(5921), 1568-1569. doi:10.1126/science.1172782

Lee, H. O., Davidson, J. M., & Duronio, R. J. (2009). Endoreplication: polyploidy with purpose. *Genes Dev, 23*(21), 2461-2477. doi:10.1101/gad.1829209

Li, Z. Y., Li, B., & Dong, A. W. (2012). The Arabidopsis transcription factor AtTCP15 regulates endoreduplication by modulating expression of key cell-cycle genes. *Mol Plant, 5*(1), 270-280. doi:10.1093/mp/ssr086

Lindemose, S., Jensen, M. K., Van de Velde, J., O'Shea, C., Heyndrickx, K. S., Workman, C. T., . . . De Masi, F. (2014). A DNA-binding-site landscape and regulatory network analysis for

NAC transcription factors in Arabidopsis thaliana. *Nucleic Acids Res, 42*(12), 7681-7693. doi:10.1093/nar/gku502

Luo, M., Cheng, K., Xu, Y., Yang, S., & Wu, K. (2017). Plant Responses to Abiotic Stress Regulated by Histone Deacetylases. *Front Plant Sci, 8*, 2147. doi:10.3389/fpls.2017.02147

Mannuss, A., Dukowic-Schulze, S., Suer, S., Hartung, F., Pacher, M., & Puchta, H. (2010). RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in Arabidopsis thaliana. *Plant Cell, 22*(10), 3318-3330. doi:10.1105/tpc.110.078568

March-Diaz, R., Garcia-Dominguez, M., Florencio, F. J., & Reyes, J. C. (2007). SEF, a new protein required for flowering repression in Arabidopsis, interacts with PIE1 and ARP6. *Plant Physiol, 143*(2), 893-901. doi:10.1104/pp.106.092270

McCord, R. A., Michishita, E., Hong, T., Berber, E., Boxer, L. D., Kusumoto, R., . . . Chua, K. F. (2009). SIRT6 stabilizes DNA-dependent protein kinase at chromatin for DNA double-strand break repair. *Aging (Albany NY), 1*(1), 109-121. doi:10.18632/aging.100011

Miki, D., Zhu, P., Zhang, W., Mao, Y., Feng, Z., Huang, H., . . . Zhu, J. K. (2017). Efficient Generation of diRNAs Requires Components in the Posttranscriptional Gene Silencing Pathway. *Sci Rep, 7*(1), 301. doi:10.1038/s41598-017-00374-7

Miller, K. M., & Jackson, S. P. (2012). Histone marks: repairing DNA breaks within the context of chromatin.

Missirian, V., Conklin, P. A., Culligan, K. M., Huefner, N. D., & Britt, A. B. (2014). High atomic weight, high-energy radiation (HZE) induces transcriptional responses shared with conventional stresses in addition to a core "DSB" response specific to clastogenic treatments. *Front Plant Sci, 5*, 364. doi:10.3389/fpls.2014.00364

Mohanta, T. K., Yadav, D., Khan, A., Hashem, A., Tabassum, B., Khan, A. L., . . . Al-Harrasi, A. (2020). Genomics, molecular and evolutionary perspective of NAC transcription factors. *Plos One, 15*(4), e0231425. doi:10.1371/journal.pone.0231425

Moldovan, G. L., & D'Andrea, A. D. (2009). How the fanconi anemia pathway guards the genome. *Annu Rev Genet, 43*, 223-249. doi:10.1146/annurev-genet-102108-134222

Morris, J. R., Boutell, C., Keppler, M., Densham, R., Weekes, D., Alamshah, A., . . . Solomon, E. (2009). The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature, 462*(7275), 886-890. doi:10.1038/nature08593

Mosammaparast, N., Kim, H., Laurent, B., Zhao, Y., Lim, H. J., Majid, M. C., . . . Shi, Y. (2013). The histone demethylase LSD1/KDM1A promotes the DNA damage response. *J Cell Biol, 203*(3), 457-470. doi:10.1083/jcb.201302092

Natale, F., Rapp, A., Yu, W., Maiser, A., Harz, H., Scholl, A., . . . Cardoso, M. C. (2017). Identification of the elementary structural units of the DNA damage response. *Nat Commun, 8*, 15760. doi:10.1038/ncomms15760

Nie, X., Wang, H., Li, J., Holec, S., & Berger, F. (2014). The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. *Biol Open, 3*(9), 794-802. doi:10.1242/bio.20148680

Nishibuchi, I., Suzuki, H., Kinomura, A., Sun, J., Liu, N. A., Horikoshi, Y., . . . Tashiro, S. (2014). Reorganization of damaged chromatin by the exchange of histone variant H2A.Z-2. *Int J Radiat Oncol Biol Phys, 89*(4), 736-744. doi:10.1016/j.ijrobp.2014.03.031

Noh, Y. S., & Amasino, R. M. (2003). PIE1, an ISWI family gene, is required for FLC activation and floral repression in Arabidopsis. *Plant Cell, 15*(7), 1671-1682. doi:10.1105/tpc.012161

O'Hagan, H. M., Mohammad, H. P., & Baylin, S. B. (2008). Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet, 4*(8), e1000155. doi:10.1371/journal.pgen.1000155

Ogita, N., Okushima, Y., Tokizawa, M., Yamamoto, Y. Y., Tanaka, M., Seki, M., . . . Umeda, M. (2018). Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in Arabidopsis. *Plant J, 94*(3), 439-453. doi:10.1111/tpj.13866

Oliveira, D. V., Kato, A., Nakamura, K., Ikura, T., Okada, M., Kobayashi, J., . . . Komatsu, K. (2014). Histone chaperone FACT regulates homologous recombination by chromatin remodeling through interaction with RNF20. *J Cell Sci, 127*(Pt 4), 763-772. doi:10.1242/jcs.135855

Oliver, C., Santos, J. L., & Pradillo, M. (2014). On the role of some ARGONAUTE proteins in meiosis and DNA repair in Arabidopsis thaliana. *Front Plant Sci, 5*, 177. doi:10.3389/fpls.2014.00177

Osakabe, K., Abe, K., Yoshioka, T., Osakabe, Y., Todoriki, S., Ichikawa, H., . . . Toki, S. (2006). Isolation and characterization of the RAD54 gene from Arabidopsis thaliana. *Plant J, 48*(6), 827-842. doi:10.1111/j.1365-313X.2006.02927.x

Ozawa, K. (2008). Reduction of phosphorylated histone H3 serine 10 and serine 28 cell cycle marker intensities after DNA damage. *Cytometry A, 73*(6), 517-527. doi:10.1002/cyto.a.20559

Panier, S., & Boulton, S. J. (2014). Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol, 15*(1), 7-18. doi:10.1038/nrm3719

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M., George, N., Fexova, S., . . . Brazma, A. (2020). Expression Atlas update: from tissues to single cells. *Nucleic Acids Res, 48*(D1), D77-D83. doi:10.1093/nar/gkz947

Pedroza-Garcia, J. A., Mazubert, C., Del Olmo, I., Bourge, M., Domenichini, S., Bounon, R., . . . Raynaud, C. (2017). Function of the Plant DNA Polymerase Epsilon in Replicative Stress Sensing, a Genetic Analysis. *Plant Physiol, 173*(3), 1735-1749. doi:10.1104/pp.17.00031

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem, 70*, 503-533. doi:10.1146/annurev.biochem.70.1.503

Preuss, S. B., & Britt, A. B. (2003). A DNA-damage-induced cell cycle checkpoint in Arabidopsis. *Genetics, 164*(1), 323-334.

Rieger, K. E., & Chu, G. (2004). Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res, 32*(16), 4786-4803. doi:10.1093/nar/gkh783

Rodgers, K., & McVey, M. (2016). Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol, 231*(1), 15-24. doi:10.1002/jcp.25053

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., & Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem, 273*(10), 5858-5868. doi:10.1074/jbc.273.10.5858

Roitinger, E., Hofer, M., Kocher, T., Pichler, P., Novatchkova, M., Yang, J., . . . Mechtler, K. (2015). Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and rad3-related (ATR) dependent DNA damage response in Arabidopsis thaliana. *Mol Cell Proteomics, 14*(3), 556-571. doi:10.1074/mcp.M114.040352

Rosa, M., Von Harder, M., Cigliano, R. A., Schlogelhofer, P., & Mittelsten Scheid, O. (2013). The Arabidopsis SWR1 chromatin-remodeling complex is important for DNA repair, somatic recombination, and meiosis. *Plant Cell, 25*(6), 1990-2001. doi:10.1105/tpc.112.104067

Rozan, L. M., & El-Deiry, W. S. (2007). p53 downstream target genes and tumor suppression: a classical view in evolution. *Cell Death Differ, 14*(1), 3-9. doi:10.1038/sj.cdd.4402058

Schalk, C., Cognat, V., Graindorge, S., Vincent, T., Voinnet, O., & Molinier, J. (2017). Small RNA-mediated repair of UV-induced DNA lesions by the DNA DAMAGE-BINDING PROTEIN 2 and ARGONAUTE 1. *Proc Natl Acad Sci U S A, 114*(14), E2965-E2974. doi:10.1073/pnas.1618834114

Schnittger, A., & De Veylder, L. (2018). The Dual Face of Cyclin B1. *Trends Plant Sci, 23*(6), 475-478. doi:10.1016/j.tplants.2018.03.015

Schonrock, N., Exner, V., Probst, A., Gruissem, W., & Hennig, L. (2006). Functional genomic analysis of CAF-1 mutants in Arabidopsis thaliana. *J Biol Chem, 281*(14), 9560-9568. doi:10.1074/jbc.M513426200

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol, 20*(11), 698-714. doi:10.1038/s41580-019-0152-0

Shaked, H., Avivi-Ragolsky, N., & Levy, A. A. (2006). Involvement of the Arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. *Genetics, 173*(2), 985-994. doi:10.1534/genetics.105.051664

Shao, H., Wang, H., & Tang, X. (2015). NAC transcription factors in plant multiple abiotic stress responses: progress and prospects. *Front Plant Sci, 6*, 902. doi:10.3389/fpls.2015.00902

Sharma, A. K., Bhattacharya, S., Khan, S. A., Khade, B., & Gupta, S. (2015). Dynamic alteration in H3 serine 10 phosphorylation is G1-phase specific during ionization radiation induced DNA damage response in human cells. *Mutat Res, 773*, 83-91. doi:10.1016/j.mrfmmm.2015.01.017

Shiloh, Y., & Ziv, Y. (2013). The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol, 14*(4), 197-210. doi:10.1038/nrm3546

Shu, Z., Row, S., & Deng, W. M. (2018). Endoreplication: The Good, the Bad, and the Ugly. *Trends Cell Biol, 28*(6), 465-474. doi:10.1016/j.tcb.2018.02.006

Sjogren, C. A., Bolaris, S. C., & Larsen, P. B. (2015). Aluminum-Dependent Terminal Differentiation of the Arabidopsis Root Tip Is Mediated through an ATR-, ALT2-, and SOG1-Regulated Transcriptional Response. *Plant Cell, 27*(9), 2501-2515. doi:10.1105/tpc.15.00172

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol, 22*(2), 96-118. doi:10.1038/s41580-020-00315-9

Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., & Jackson, S. P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell, 123*(7), 1213-1226. doi:10.1016/j.cell.2005.09.038

Su, H., Cheng, Z., Huang, J., Lin, J., Copenhaver, G. P., Ma, H., & Wang, Y. (2017). Arabidopsis RAD51, RAD51C and XRCC3 proteins form a complex and facilitate RAD51 localization on chromosomes for meiotic recombination. *PLoS Genet, 13*(5), e1006827. doi:10.1371/journal.pgen.1006827

Su, T. T. (2006). Cellular responses to DNA damage: one signal, multiple choices. *Annu Rev Genet, 40*, 187-208. doi:10.1146/annurev.genet.40.110405.090428

Sundaramoorthy, R., & Owen-Hughes, T. (2020). Chromatin remodelling comes into focus. *F1000Res, 9*. doi:10.12688/f1000research.21933.1

Sunkar, R., Kapoor, A., & Zhu, J. K. (2006). Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell, 18*(8), 2051-2065. doi:10.1105/tpc.106.041673

Takahashi, N., & Umeda, M. (2014). Cytokinins promote onset of endoreplication by controlling cell cycle machinery. *Plant Signal Behav, 9*(8), e29396. doi:10.4161/psb.29396

Tamura, K., Adachi, Y., Chiba, K., Oguchi, K., & Takahashi, H. (2002). Identification of Ku70 and Ku80 homologues in Arabidopsis thaliana: evidence for a role in the repair of DNA double-strand breaks. *Plant J, 29*(6), 771-781. doi:10.1046/j.1365-313x.2002.01258.x

Tjeertes, J. V., Miller, K. M., & Jackson, S. P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J, 28*(13), 1878-1889. doi:10.1038/emboj.2009.119

Turinetto, V., & Giachino, C. (2015). Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Res, 43*(5), 2489-2498. doi:10.1093/nar/gkv061

Uckelmann, M., Densham, R. M., Baas, R., Winterwerp, H. H. K., Fish, A., Sixma, T. K., & Morris, J. R. (2018). USP48 restrains resection by site-specific cleavage of the BRCA1 ubiquitin mark from H2A. *Nat Commun, 9*(1), 229. doi:10.1038/s41467-017-02653-3

Valdes-Mora, F., Song, J. Z., Statham, A. L., Strbenac, D., Robinson, M. D., Nair, S. S., . . . Clark, S. J. (2012). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res, 22*(2), 307-321. doi:10.1101/gr.118919.110

van Attikum, H., Bundock, P., Overmeer, R. M., Lee, L. Y., Gelvin, S. B., & Hooykaas, P. J. (2003). The Arabidopsis AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. *Nucleic Acids Res, 31*(14), 4247-4255. doi:10.1093/nar/gkg458

Varas, J., Santos, J. L., & Pradillo, M. (2017). The Absence of the Arabidopsis Chaperone Complex CAF-1 Produces Mitotic Chromosome Abnormalities and Changes in the Expression Profiles of Genes Involved in DNA Repair. *Front Plant Sci, 8*, 525. doi:10.3389/fpls.2017.00525

Vlieghe, K., Boudolf, V., Beemster, G. T., Maes, S., Magyar, Z., Atanassova, A., . . . De Veylder, L. (2005). The DP-E2F-like gene DEL1 controls the endocycle in Arabidopsis thaliana. *Curr Biol, 15*(1), 59-63. doi:10.1016/j.cub.2004.12.038

Waititu, J. K., Zhang, C., Liu, J., & Wang, H. (2020). Plant Non-Coding RNAs: Origin, Biogenesis, Mode of Action and Their Roles in Abiotic Stress. *Int J Mol Sci, 21*(21). doi:10.3390/ijms21218401

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature, 431*(7010), 873-878. doi:10.1038/nature02985

Wang, S., Durrant, W. E., Song, J., Spivey, N. W., & Dong, X. (2010). *Arabidopsis* BRCA2 and RAD51 proteins are specifically involved in defense gene transcription during plant immune responses. *Proc Natl Acad Sci U S A, 107*(52), 22716-22721. doi:10.1073/pnas.1005978107

Wang, Y., Xiao, R., Wang, H., Cheng, Z., Li, W., Zhu, G., . . . Ma, H. (2014). The Arabidopsis RAD51 paralogs RAD51B, RAD51D and XRCC2 play partially redundant roles in somatic DNA repair and gene regulation. *New Phytol, 201*(1), 292-304. doi:10.1111/nph.12498

Wang, Z., Schwacke, R., & Kunze, R. (2016). DNA Damage-Induced Transcription of Transposable Elements and Long Non-coding RNAs in Arabidopsis Is Rare and ATM-Dependent. *Mol Plant, 9*(8), 1142-1155. doi:10.1016/j.molp.2016.04.015

Waterworth, W. M., Footitt, S., Bray, C. M., Finch-Savage, W. E., & West, C. E. (2016). DNA damage checkpoint kinase ATM regulates germination and maintains genome stability in seeds. *Proc Natl Acad Sci U S A, 113*(34), 9647-9652. doi:10.1073/pnas.1608829113

Weber, C. M., Ramachandran, S., & Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell, 53*(5), 819-830. doi:10.1016/j.molcel.2014.02.014

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., . . . Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. *Cell, 149*(1), 101-112. doi:10.1016/j.cell.2012.03.002

Weimer, A. K., Biedermann, S., Harashima, H., Roodbarkelari, F., Takahashi, N., Foreman, J., . . . Schnittger, A. (2016). The plant-specific CDKB1-CYCB1 complex mediates homologous recombination repair in Arabidopsis. *EMBO J, 35*(19), 2068-2086. doi:10.15252/embj.201593083

West, C. E., Waterworth, W. M., Jiang, Q., & Bray, C. M. (2000). Arabidopsis DNA ligase IV is induced by gamma-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. *Plant J, 24*(1), 67-78. doi:10.1046/j.1365-313x.2000.00856.x

Wood, R. D. (2020). Human DNA Repair Genes. Retrieved from https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html

Wu, L., Liu, S., Qi, H., Cai, H., & Xu, M. (2020). Research Progress on Plant Long Non-Coding RNA. *Plants (Basel), 9*(4), 408. doi:10.3390/plants9040408

Xu, P., Yuan, D., Liu, M., Li, C., Liu, Y., Zhang, S., . . . Yang, C. (2013). AtMMS21, an SMC5/6 complex subunit, is involved in stem cell niche maintenance and DNA damage responses in Arabidopsis roots. *Plant Physiol, 161*(4), 1755-1768. doi:10.1104/pp.112.208942

Xu, Y., Ayrapetov, M. K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y., & Price, B. D. (2012). Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. *Mol Cell, 48*(5), 723-733. doi:10.1016/j.molcel.2012.09.026

Yi, D., Alvim Kamei, C. L., Cools, T., Vanderauwera, S., Takahashi, N., Okushima, Y., . . . De Veylder, L. (2014). The Arabidopsis SIAMESE-RELATED cyclin-dependent kinase inhibitors SMR5 and SMR7 regulate the DNA damage checkpoint in response to reactive oxygen species. *Plant Cell, 26*(1), 296-309. doi:10.1105/tpc.113.118943

Yoshiyama, K., Conklin, P. A., Huefner, N. D., & Britt, A. B. (2009). Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A, 106*(31), 12843-12848. doi:10.1073/pnas.0810304106

Yoshiyama, K. O. (2016). SOG1: a master regulator of the DNA damage response in plants. *Genes Genet Syst, 90*(4), 209-216. doi:10.1266/ggs.15-00011

Yoshiyama, K. O., Kaminoyama, K., Sakamoto, T., & Kimura, S. (2017). Increased Phosphorylation of Ser-Gln Sites on SUPPRESSOR OF GAMMA RESPONSE1 Strengthens the DNA Damage Response in Arabidopsis thaliana. *Plant Cell, 29*(12), 3255-3268. doi:10.1105/tpc.17.00267

Yoshiyama, K. O., Kimura, S., Maki, H., Britt, A. B., & Umeda, M. (2014). The role of SOG1, a plant-specific transcriptional regulator, in the DNA damage response. *Plant Signal Behav, 9*(4), e28889. doi:10.4161/psb.28889

Yoshiyama, K. O., Kobayashi, J., Ogita, N., Ueda, M., Kimura, S., Maki, H., & Umeda, M. (2013). ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in Arabidopsis. *EMBO Rep, 14*(9), 817-822. doi:10.1038/embor.2013.112

Yoshiyama, K. O., Sakaguchi, K., & Kimura, S. (2013). DNA damage response in plants: conserved and variable response compared to animals. *Biology (Basel), 2*(4), 1338-1356. doi:10.3390/biology2041338

Yuan, D., Lai, J., Xu, P., Zhang, S., Zhang, J., Li, C., . . . Yang, C. (2014). AtMMS21 regulates DNA damage response and homologous recombination repair in Arabidopsis. *DNA Repair (Amst), 21*, 140-147. doi:10.1016/j.dnarep.2014.04.006

Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., . . . Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell, 153*(1), 193-205. doi:10.1016/j.cell.2013.02.033

Zhang, C., Cao, L., Rong, L., An, Z., Zhou, W., Ma, J., . . . Dong, A. (2015). The chromatin-remodeling factor AtINO80 plays crucial roles in genome stability maintenance and in plant development. *Plant J, 82*(4), 655-668. doi:10.1111/tpj.12840

Zhang, L. N., Li, J. Y., & Xu, W. (2013). A review of the role of Puma, Noxa and Bim in the tumorigenesis, therapy and drug resistance of chronic lymphocytic leukemia. *Cancer Gene Ther, 20*(1), 1-7. doi:10.1038/cgt.2012.84

Zhao, D., Li, Y., Xiong, X., Chen, Z., & Li, H. (2017). YEATS Domain-A Histone Acylation Reader in Health and Disease. *J Mol Biol, 429*(13), 1994-2002. doi:10.1016/j.jmb.2017.03.010

Zhao, S., Zhang, B., Yang, M., Zhu, J., & Li, H. (2018). Systematic Profiling of Histone Readers in Arabidopsis thaliana. *Cell Rep, 22*(4), 1090-1102. doi:10.1016/j.celrep.2017.12.099

Zhou, K., Gaullier, G., & Luger, K. (2019). Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol, 26*(1), 3-13. doi:10.1038/s41594-018-0166-x

Zhou, X., Wang, G., & Zhang, W. (2007). UV-B responsive microRNA genes in Arabidopsis thaliana. *Mol Syst Biol, 3,* 103. doi:10.1038/msb4100143

Zhu, Y., Weng, M., Yang, Y., Zhang, C., Li, Z., Shen, W. H., & Dong, A. (2011). Arabidopsis homologues of the histone chaperone ASF1 are crucial for chromatin replication and cell proliferation in plant development. *Plant J, 66*(3), 443-455. doi:10.1111/j.1365-313X.2011.04504.x

Chapter 2 UNDERSTANDING THE TRANSCRIPTIONAL COMPONENT OF THE DNA DAMAGE RESPONSE

CHAPTER SUMMARY

The DNA damage response (DDR) in *Arabidopsis* involves a massive transcriptional response, mediated by the SOG1 transcription factor, that regulates the cell cycle, DNA repair, and cell death. Despite the central role of SOG1 in the DNA damage response, and past efforts showing that SOG1 is critical for coping with DNA damage, the expression networks downstream of SOG1, which are required to regulate key DDR processes (*e.g.*, cell cycle, DNA repair and programmed cell death), remain largely unknown. Moreover, global expression defects in wild-type and *sog1* mutants have only been assessed at single time points after DNA damage and thus were not able to capture the temporal transcriptional DNA damage response, specifically the very early and very late transcriptional responses. Thus, the expression dynamics of the DNA damage response, the full extent of SOG1's role in gene regulation, and the transcriptional networks linking SOG1 to specific damage-associated processes were yet to be determined. These gaps were addressed in the form of a PNAS paper on which I was a second author (contributed to experimental design, execution, and analysis of the mRNAseq and ChIPseq data), in section "**SOG1 and MYB3R regulate the DNA damage network in *Arabidopsis*".** This work allowed the identification of the transcriptional dynamics of the various DDR processes and identified SOG1 as the major activator of the transcriptional response to DNA damage, directly targeting ~300 genes, and MYB3R as the major repressors of cell cycle genes in the DNA damage response. While SOG1 has been equated to the master regulator of the DNA damage response, based on prior studies, this work revealed SOG1-independent aspects of the DNA damage response whose regulation is yet to be deciphered. This transcriptional roadmap provides insight into the regulators

and dynamic processes coordinated in response to gamma-irradiation and serves as a framework for the generation of additional hypothesis to expand our understanding of the DDR.

SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis*

Clara Bourbousse[a,1], Neeraja Vegesna[a,b], and Julie A. Law[a,b,2]

[a]Plant Molecular and Cellular Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037; and [b]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093

Edited by Julia Bailey-Serres, University of California, Riverside, CA, and approved November 14, 2018 (received for review June 21, 2018)

To combat DNA damage, organisms mount a DNA damage response (DDR) that results in cell cycle regulation, DNA repair and, in severe cases, cell death. Underscoring the importance of gene regulation in this response, studies in *Arabidopsis* have demonstrated that all of the aforementioned processes rely on SUPPRESSOR OF GAMMA RESPONSE 1 (SOG1), a NAC family transcription factor (TF) that has been functionally equated to the mammalian tumor suppressor, p53. However, the expression networks connecting SOG1 to these processes remain largely unknown and, although the DDR spans from minutes to hours, most transcriptomic data correspond to single time-point snapshots. Here, we generated transcriptional models of the DDR from GAMMA (γ)-irradiated wild-type and *sog1* seedlings during a 24-hour time course using DREM, the Dynamic Regulatory Events Miner, revealing 11 coexpressed gene groups with distinct biological functions and *cis*-regulatory features. Within these networks, additional chromatin immunoprecipitation and transcriptomic experiments revealed that SOG1 is the major activator, directly targeting the most strongly up-regulated genes, including TFs, repair factors, and early cell cycle regulators, while three MYB3R TFs are the major repressors, specifically targeting the most strongly down-regulated genes, which mainly correspond to G2/M cell cycle-regulated genes. Together these models reveal the temporal dynamics of the transcriptional events triggered by γ-irradiation and connects these events to TFs and biological processes over a time scale commensurate with key processes coordinated in response to DNA damage, greatly expanding our understanding of the DDR.

DNA damage response | DREM | SOG1 | transcriptional networks

The genomes of all organisms incur various types of DNA damage due to both endogenous processes and exposure to exogenous stresses or toxic compounds (1, 2). Of this damage, DNA double-strand breaks (DSBs) are particularly hazardous, as no intact strand remains to guide the DNA repair, potentially leading to chromosomal deletions and translocations (3, 4). To cope with such damage, mechanisms are in place to sense DNA lesions and initiate a DNA damage response (1, 5). This response involves the transcriptional and posttranscriptional regulation of diverse cellular pathways, ultimately leading to DNA repair, via the expression and/or targeting of repair factors to sites of damage, to cell cycle arrest, which provides additional time for DNA repair before replication, or to cell death, when the damage is too severe (5, 6). Given the importance of maintaining genome stability for proper cellular function and the faithful inheritance of genetic information (1–3, 5), it is critical to understand how the DNA damage response is initiated, coordinated, and executed.

Studies in yeast, plants, and mammals have revealed many highly conserved aspects of the DNA damage response (6–8). In the case of DSBs, conserved sensors, namely the MRN and Ku70/80 complexes, recognize the damaged DNA and transducers, including the ATAXIA-TELANGIECTASIA MUTATED (ATM) and ATAXIA-TELANGIECTASIA MUTATED AND RAD3-RELATED (ATR) kinases, initiate signaling cascades through the posttranslational modification of target proteins (1, 8–10). These cascades modulate the activities of both shared and organism-specific effector proteins, culminating in the regulation of DSB repair, via homologous recombination and various nonhomologous end joining pathways, as well as the regulation of gene expression, cell cycle arrest, cell death, and endoreduplication (1, 6, 8, 11). To gain insight into the pathways and molecular interactions orchestrating these events, efforts in many organisms have focused on identifying and characterizing the key players, signaling cascades, and transcriptional programs that stem from the recognition of DNA damage.

In plants, the SUPPRESSOR OF GAMMA RESPONSE 1 (SOG1) transcription factor (TF) was identified from a DNA damage-suppressor screen (12) and was shown to be a major regulator of the DNA damage response (13). In the absence of SOG1, *Arabidopsis* plants exposed to DNA damaging agents display defects in gene regulation (13), cell cycle arrest (12), programmed cell death (14), endoreduplication (15), DNA repair, and genome stability (12, 13). These findings, along with those showing that SOG1 is regulated in an ATM-dependent manner via phosphorylation of conserved serine-glutamine motifs (16, 17), have led to SOG1 being functionally equated with p53 (8, 18), a mammalian tumor suppressor that coordinates the DNA damage response and is also phosphorylated in an ATM/ATR-dependent manner (19, 20).

Despite the central role of SOG1 in the DNA damage response, and the numerous studies showing SOG1 is critical for coping with DNA damage (12–15, 21–26), global expression defects in *sog1* mutants have only been assessed at single time points following γ-irradiation (γ-IR) (2 h) (13) or zeocin (1.5 h) (27) and, until recently (27), only a few SOG1 targets had been identified (22, 25, 26,

Significance

DNA damage triggers a highly conserved response that coordinates processes necessary to maintain genome integrity, including cell cycle arrest, DNA repair, and cell death. Despite the identification of primary transcription factors (TFs) that control these processes, knowledge regarding the downstream genes and regulatory networks controlled by these TFs remains poorly understood. Using *Arabidopsis*, we generated the first model of the DNA damage response transcriptional network, revealing 11 coexpressed gene groups with distinct biological functions and *cis*-regulatory features. Our characterization of this model demonstrates that SOG1 and three MYB3R TFs are, respectively, the major activator and repressors within this network, coordinating the rapid induction of DNA repair genes and TF cascades as well as the subsequent repression of cell cycle genes.

Author contributions: C.B., N.V., and J.A.L. designed research; C.B. and N.V. performed research; C.B. and J.A.L. analyzed data; and C.B. and J.A.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The source data files and sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE112773).

[1]Present address: Institut de biologie de l'Ecole normale supérieure, Ecole normale supérieure, CNRS, INSERM, Paris Sciences et Lettres Université, 75005 Paris, France.

[2]To whom correspondence should be addressed. Email: jlaw@salk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810582115/-/DCSupplemental.

Published online December 12, 2018.

28). Furthermore, although the perception of DNA damage caused by exposure to γ-IR triggers events that occur on a time scale of minutes [e.g., the ATM/ATR-dependent phosphorylation of H2AX at DSBs (29–31)] to hours [e.g., cell cycle regulation (12, 29)], our understanding of the transcriptional changes coordinating these events is largely restricted to profiling experiments conducted at discrete time points (13, 32–39). Extending on these transcriptional snapshots, two previous studies profiled gene expression across several time points, but they utilized early array technology (40) or only included controls at a subset of time points (41). Thus, the expression dynamics of the DNA damage response, the full extent of SOG1's role in gene regulation, and the transcriptional networks linking SOG1 to specific damage-associated processes remain to be determined.

To reveal the temporal features of the transcriptional response to DNA damage, and to further investigate the roles of SOG1 in executing this response, we performed transcriptomic analyses using γ-IR–treated wild-type and *sog1* seedlings over a 24-h time course. These data, along with literature-curated gene–TF interactions, were then used to generate transcriptional network models of the *Arabidopsis* DNA damage response via DREM, the Dynamic Regulator Events Miner (42, 43). In total, ∼2,400 differentially expressed (DE) genes were identified, greatly expanding upon the previously identified DNA damage-responsive genes. In the wild-type DREM model, these genes were organized into 11 coexpressed groups with distinct expression profiles, promoter motifs, and gene ontology (GO) enrichments. Using this DREM model as a guide, additional analyses revealed both SOG1-dependent and -independent aspects of the DNA damage response and demonstrated that in addition to controlling the induction of many γ-IR responsive genes, SOG1 is also required for the repression of hundreds of genes. Furthermore, despite this dual effect in gene regulation, we found that SOG1 acts exclusively as a transcriptional activator, directly targeting ∼300 genes, including many DNA repair and cell cycle factors, as well as a large subset of TFs, placing it at the top of a complex gene regulatory network. Finally, gene-expression analysis of the *myb3r1,3,5* triple mutant revealed that these TFs repress a large subset of G2/M-specific genes in response to DNA damage. Taken together, our findings not only shed light on the DNA damage response, but also provide a framework to begin connecting specific expression subnetworks to the diverse biological processes coordinated during this response.

Results and Discussion

Temporal Characterization of the DNA Damage Response Reveals Coexpressed Gene Sets with Distinct Biological Functions and Regulatory Features. To obtain a temporal view of the expression networks underpinning the DNA damage response in *Arabidopsis*, mRNA sequencing (mRNA-seq) experiments were conducted at six time points from 20 min to 24 h after either mock or γ-IR treatments in wild-type plants (*SI Appendix*, Fig. S1*A* and Dataset S1). Furthermore, as the SOG1 TF is known to regulate many genes induced by DNA damage (13), a γ-IR time course experiment was also conducted in the *sog1* mutant (*SI Appendix*, Fig. S1*A* and Dataset S1). Consistent with having selected a suitable time scale to capture the dynamics of the DNA damage response, identification of DE genes during the wild-type γ-IR time course (Dataset S2 *A* and *C*) [fold-change (FC) ≥ 2, false-discovery rate (FDR) ≤ 0.01] revealed that of the ∼2,400 DE genes identified, 320 were already observed after just 20 min, while only 141 remained by 24 h (*SI Appendix*, Fig. S1*B*).

To determine whether the gene-expression changes observed in our γ-IR time course are consistent with previous assessments of the DNA damage response, the DE genes identified in at least one time point from the wild-type time course were compared with four previously published DE gene sets identified using similar γ-IR treatments and seedling stages, but at single time points (13, 32, 33, 38) (*SI Appendix*, Fig. S1 *C* and *D*). These analyses revealed that our γ-IR time course captures the vast majority of previously identified up-regulated genes, 65.2–91.4%, with the three datasets based on the most similar conditions and stages (13, 32, 33) showing the greatest overlaps (*SI Appendix*, Fig. S1*C*). For down-regulated genes, the overlap was lower but many common genes were still identified (*SI Appendix*, Fig. S1*D*). Based on these overlaps, we conclude that our γ-IR time course represents a stereotypical DNA damage response. However, in addition to the previously identified DE genes, over half of the up- and down-regulated genes identified in the γ-IR time course (655 and 519, respectively) were unique to this study (*SI Appendix*, Fig. S1 *C* and *D*). Approximately 60% of these uniquely identified genes were regulated at commonly assessed time points (i.e., 1 h 30 min or 3 h) (*SI Appendix*, Fig. S1*E*), which likely reflects the increased sensitivity imparted by the use of mRNA sequencing rather than microarrays. The remaining 40% were specific to the additional time points covered in the γ-IR time course (*SI Appendix*, Fig. S1*E*). Collectively, these analyses identified an expanded set of DNA damage-responsive genes over a time span that corresponds to major events occurring in response to DNA damage, thereby providing a more comprehensive view of the DNA damage response.

To gain insight into how the thousands of γ-IR–regulated genes coordinate the DNA damage response, we used the DREM software (42, 43) to generate a temporal model of the underlying gene-regulatory network. This software groups genes with similar expression patterns together and assigns putative TFs to specific gene sets based on previously identified gene–TF interactions, enabling predictions to be made regarding specific features of the network. To facilitate subsequent comparisons with data generated in the *sog1* mutant, the wild-type DREM model was constructed based on the $\log_2$ FC in expression (γ-IR vs. mock-treated) of the 2,395 DE genes [2,177 DE genes (FC ≥ 2 and FDR ≤ 0.01) from the wild-type γ-IR time course plus 218 additional DE genes specific to the *sog1* γ-IR time course] [*SI Appendix*, Fig. S6*B*, Dataset S2*B*, and Source Data 1 (see ref. 44 for all source data files)]. For the gene–TF interactions, data were compiled from the AGRIS (Arabidopsis Gene Regulatory Information Server) (45–47) and DAP-seq databases (48). As these data represent a mixture of in vitro- and in vivo-derived interactions, the DREM model was generated based solely on the expression data, with the TF predictions determined subsequently. The resulting wild-type DREM model revealed 11 groups of coexpressed genes with distinct biological functions and regulatory features (Fig. 1).

At the level of gene expression, investigation of the DREM network confirmed that the 11 coexpressed groups, designated as paths W1–W11 (Fig. 1*A* and Dataset S3*A*), display highly correlated expression profiles across the entire γ-IR time course (*SI Appendix*, Fig. S2*A*) and capture a wide range of expression dynamics, as exemplified by the representative gene-expression patterns shown in Fig. 1*B*. The two most prominent features revealed by the DREM analysis are a subset of paths (W1–W4) showing broad peaks of induction around 3 h and another subset (W9–W11) showing broad peaks of repression between 3 and 6 h. The remaining paths correspond to early up-regulated genes (W5 at 1 h 30 min and W6 at 20 min), early down-regulated genes (W8 at 1 h 30 min), and late mildly up-regulated genes (W7 at 3–12 h). Notably, ∼half of the genes in the early and late responsive paths (W5–W8), as well as many of the genes in the other paths, were uniquely identified in our γ-IR time course (*SI Appendix*, Fig. S2*B*). As expected, the 218 DE genes included based solely on the *sog1* γ-IR time course displayed no significant changes in expression in the wild-type DREM model (*SI Appendix*, Fig. S2*C*). Thus, while many genes show peaks of induction or repression at commonly assessed time points (1 h 30 min to 3 h), this DREM model reveals additional expression modules that peak earlier or later, providing insights into the DNA damage response.

To shed light on the biological functions of these gene sets, GO analyses were performed, revealing largely distinct enrichment terms for the DREM paths that capture the major processes previously connected with the DNA damage response [Fig. 1*C*, *SI Appendix*, Fig. S3, and Source Data 2 (44)]. These include DNA repair and DNA metabolism terms for genes up-regulated in paths W1, W2, and—to a lesser extent—W3, cell cycle and associated terms for genes down-regulated in paths W9–W11, cell death terms for genes in paths W4 and W6, and respiratory burst along with other reactive oxygen species-associated terms for paths W6 and W7 (Fig. 1 *C* and *D* and *SI*

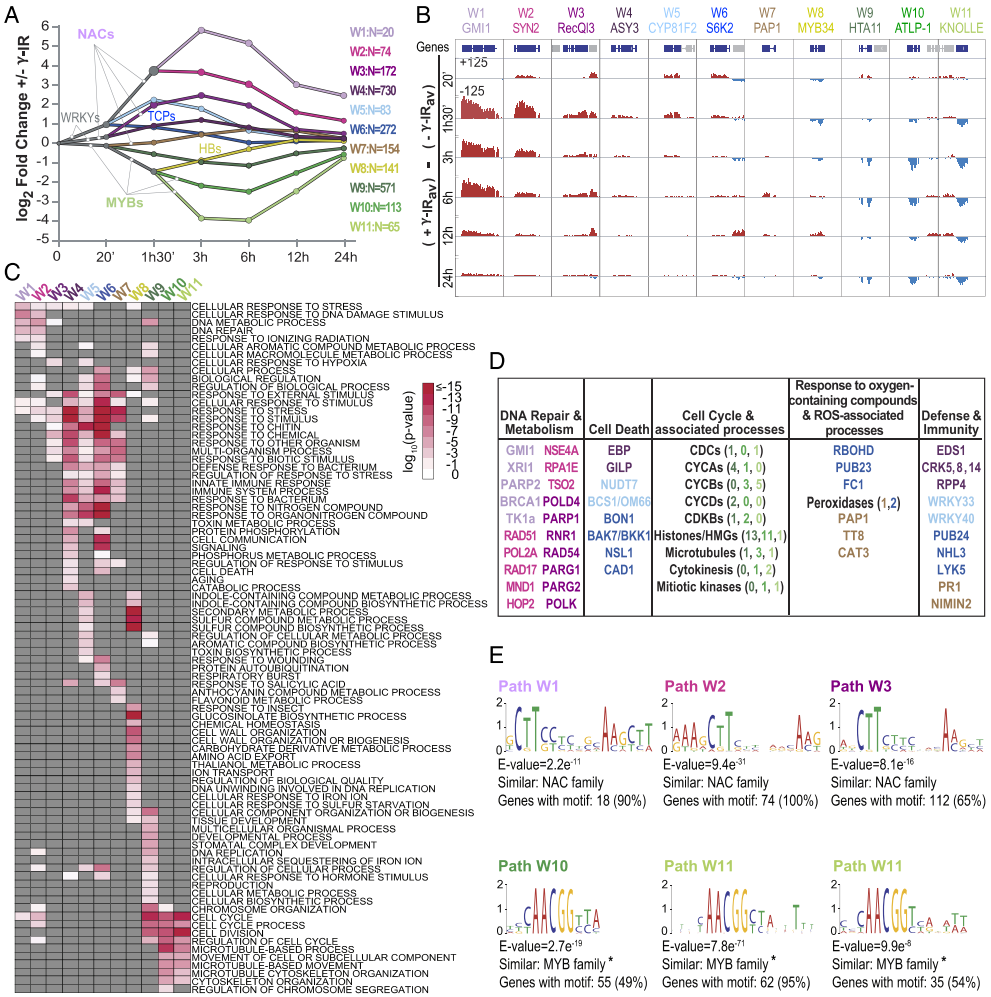


Fig. 1. DNA damage response DREM analysis reveals coexpressed genes with distinct biological functions and regulatory features. (*A*) DREM model [see Source Data 1 (44)] showing 11 groups of coexpressed genes, termed wild-type paths W1–W11. Here, and in all other DREM models, the *y* axis indicates the $\log_2$ FC in expression in response to γ-IR, the *x* axis indicates the time in minutes (') and/or hours (h), and the number (N) of genes per path is indicated. All genes are listed in Dataset S3A. Comparisons with previously published DE gene sets are presented in *SI Appendix*, Fig. S1, expression patterns of the individual genes in each DREM path are shown in *SI Appendix*, Fig. S2*A*, and the TF families (i.e., NAC, TCP, HB, WRKY, and MYB) assigned to the DREM paths are indicated, with the lists of all of the TFs assigned to each path shown in *SI Appendix*, Fig. S4. (*B*) Screenshots showing the expression levels of representative genes from each DREM path. The gene indicated above is shown in blue and the neighboring genes are shown in gray. The difference between the mock and γ-IR–treated samples $[(+\gamma\text{-IR}_{av}) - (-\gamma\text{-IR}_{av})]$ is shown for each time point on a scale of ±125. (*C*) Heatmap showing the significance of all of the GO terms with a $\log_{10} P < -1.7$ in at least one path, across all of the DREM paths. Gray indicates a $P > 0.5$. See *SI Appendix*, Fig. S3 and Source Data 2 (44) for all enriched GO terms remaining after the REVIGO similarity filter (100). (*D*) Table of select genes associated with the processes indicated above each column based on GO category enrichments and/or manual curation. The gene names and/or the number of genes per category (in parentheses) are colored based on the path in which they reside. (*E*) Select motifs enriched in the indicated DREM paths. The full set of enriched motifs identified are presented in *SI Appendix*, Fig. S5 [Source Data 3 (44)]. Below each motif, E-values calculated against all *Arabidopsis* promoter sequences, and similar TF families, identified via comparisons with the motifs in the DAP-seq database (48) using Tomtom (98), are indicated. Three similar, but independently identified motifs that correspond to the previously described mitosis-specific activator (MSA) element (AACGG) (55) are marked with an asterisk.

Appendix, Fig. S3). In addition, several paths are enriched for immunity and defense terms (Fig. 1 *C* and *D* and *SI Appendix*, Fig. S3), which is consistent with previous work showing both transcriptional and genetic connections between the plant immunity and DNA damage responses (49, 50). Thus, although many of the genes in the DREM model remain poorly characterized (Dataset S3*A*), grouping the various DNA damage-responsive genes into paths enabled the identification of pertinent GO terms. Furthermore, these analyses revealed that, broadly speaking, the transcriptional response to DNA damage starts with a transient induction of general stress genes, which is coincident with the sustained induction of DNA repair genes, and is followed, after a short delay, by the repression of cell cycle genes. Together, these findings add an important layer of functional and temporal information to the DNA-damage gene-regulatory network.

Finally, to better understand how these genes (and associated biological processes) are controlled, putative TFs associated with the DNA damage response were identified using a two-pronged approach. Specifically, the DREM analysis, which leverages previously identified gene–TF interactions, was complimented by motif search analyses, which are unbiased and thus have the potential to reveal novel gene–TF interactions. Both of these approaches suggest major roles for the NAC and MYB3R TF families in the gene regulatory network and implicate more minor roles for several other TF families in connection with specific DREM paths (Fig. 1*A* and *SI Appendix*, Fig. S4). For the NAC and MYB TF assignments, the families, although not necessarily all of the specific TFs, identified by the DREM and motif analyses are consistent with current knowledge of the DNA damage response. For example, the three most strongly up-regulated paths (W1–W3) were assigned to numerous NAC TFs (Fig. 1*A* and *SI Appendix*, Fig. S4), and known NAC motifs (CTT[N_7]AAG) (48, 51, 52) were identified in the promoters of a high percentage of the genes in these paths [Fig. 1*E*, *SI Appendix*, Fig. S5, and Source Data 3 (44)]. Thus, although SOG1 (also known as NAC008) was not assigned to these paths because there was no SOG1–gene interaction data available in the AGRIS (45–47) and DAP-seq (48) databases, the NAC family was well represented. Similarly, all members of the MYB3R TF family present in the DAP-seq database (MYB3R1, -4, and -5) were assigned to the two most strongly down-regulated paths in the DREM model (W10 and W11) (Fig. 1*A* and *SI Appendix*, Fig. S4), but only the repressor MYB3Rs (Rep-MYB3Rs), MYB3R3 and MYB3R5, have known roles in the DNA damage response (53). Consistent with these MYB3R TF assignments, and the GO enrichment terms associated with paths W10 and W11, an AACGG motif that is bound by several MYB3R TFs (54), and was originally described as the mitosis-specific activator (MSA) motif based on its enrichment in the promoters of G2/M-specific genes (55–57), was identified in the promoters of the path W10 and W11 genes (Fig. 1*E* and *SI Appendix*, Fig. S5). Based on these TF predictions, additional genetic and genomic analyses were conducted to assign SOG1 and the Rep-MYB3R TFs to the regulation of specific genes and reveal mechanistic insights into how these TFs coordinate the DNA damage response.

SOG1 Controls Nearly All Transcriptional Aspects of the DNA Damage Response. Previously, SOG1 was shown to control the expression of ∼300 up-regulated, and a few down-regulated genes, 1.5 h after exposure to γ-IR, demonstrating it plays a major role in gene regulation during the DNA damage response (13). Here, many additional DNA damage-responsive genes were identified as part of the wild-type γ-IR time course. To determine the role of SOG1 in regulating this network, a new DREM model was computed based on the same 2,395 genes included in the wild-type analysis (2,177 DE genes from the wild-type γ-IR time course and 218 *sog1*-specific DE genes), but using the RNA-seq data from the *sog1* γ-IR time course. In this model, only five paths were identified, and they displayed a striking reduction in their expression changes compared with the wild-type DREM paths (Fig. 2*A*, *SI Appendix*, Fig. S6*A*, and Dataset S3*B*). Consistent with the identification of fewer DE genes in the *sog1* mutant (771; FC ≥ 2 and FDR ≤ 0.01), most of the genes included in the *sog1* DREM model based solely on their regulation in the wild-type γ-IR time course (1,451 of 1,624) reside in paths S3 and S4 and showed little to no concerted changes in expression across the entire *sog1* γ-IR time course (*SI Appendix*, Fig. S6 *B* and *C*). The remaining 173 genes, which did not pass the thresholds for differential expression in the *sog1* γ-IR time course, displayed subtle, but more concerted, expression changes and were assigned to paths S1, S2, or S5 in the *sog1* DREM model (*SI Appendix*, Fig. S6*C*). Thus, nearly two-thirds (1,451 of 2,395) of the transcriptional DNA damage response is completely dependent on SOG1. As detailed below, assessment of the features remaining in the *sog1* DREM network (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S6*D*) and visualization of the *sog1* expression patterns within the context of the wild-type DREM model (Fig. 2*C*) reveal both SOG1-dependent and SOG1-independent aspects of the DNA damage response.

Analysis of the path S1 genes from the *sog1* DREM model revealed substantial overlaps with the genes present in paths W5 and W6 of the wild-type DREM model (Fig. 2*B*). Furthermore, the expression levels of many of the W5 and W6 genes were comparable between the wild-type and *sog1* datasets at the 20-min time point (Fig. 2*C*), demonstrating that a large portion of the initial burst of gene expression observed in the wild-type DNA damage response occurs in a SOG1-independent manner. However, for the path W5 genes, the sustained induction observed in the wild-type γ-IR time course was dependent on SOG1 (Fig. 2*C*). Importantly, the up-regulation of path W5 and W6 genes 20-min post-irradiation and the SOG1-independent nature of this early response were independently verified by a second set of wild-type and *sog1* γ-IR time courses (*SI Appendix*, Fig. S7). Because many of the genes in path S1, like those in paths W5 and W6, are associated with diverse stress-response terms (Fig. 1*C* and *SI Appendix*, Figs. S6*D* and S8*A*), and because these expression profiles show similarities to profiles observed after various biotic and abiotic stresses (*SI Appendix*, Fig. S8*B*) rather than being quite specific for genotoxic stress, like W1–W3, we posit that this SOG1-independent aspect of the DNA damage response likely represents a more general stress response that may not be directly coupled to the detection of damaged DNA.

Analysis of the path S5 genes from the *sog1* DREM model revealed that a specific subset of genes are repressed in a partially SOG1-independent manner (Fig. 2*A*). These genes correspond almost exclusively (98%) to paths W10 and W11 of the wild-type DREM model (Fig. 2*B*) and, consistent with this high degree of overlap, they are enriched for cell cycle-associated genes and show similar promoter motifs as those observed for paths W10 and W11, including the MYB/MSA motif (*SI Appendix*, Figs. S6*D* and S9). Strikingly, these path S5 genes include 92% of the genes in path W11 (Fig. 2*B*), demonstrating that these genes are still repressed, although to a lower extent, in *sog1* mutants. Indeed, comparisons of the gene-expression profiles across all of the down-regulated paths in the wild-type DREM model demonstrate that the genes present in paths W8 and W9 are strongly SOG1-dependent, while those in paths W10 and W11 are only partially SOG1-dependent (Fig. 2*C*). This partial dependence on SOG1, and selectivity for the path W10 and W11 genes, was independently verified by a second set of wild-type and *sog1* γ-IR time courses (*SI Appendix*, Fig. S7). Finally, these findings are also in agreement with published qRT-PCR data showing that the suppression of two cell-cycle genes in response to γ-IR (*CDKB2;1* and *KNOLLE*, which are present in paths W10 and W11, respectively) are only partially SOG1-dependent (13). Together, these analyses reveal a specific subset of strongly repressed cell cycle genes that are regulated by both SOG1-dependent and SOG1-independent pathways during the DNA damage response.

Analysis of the path S2 genes from the *sog1* DREM model revealed a latent DNA damage response that is prominent in the *sog1* mutant 24 h after irradiation (Fig. 2 *A* and *C* and *SI Appendix*, Fig. S6*A*). Interestingly, two-thirds of these genes correspond to those found in paths W1–W4 of the wild-type DREM model (Fig. 2*B*), which normally peak between 1 h 30 min and

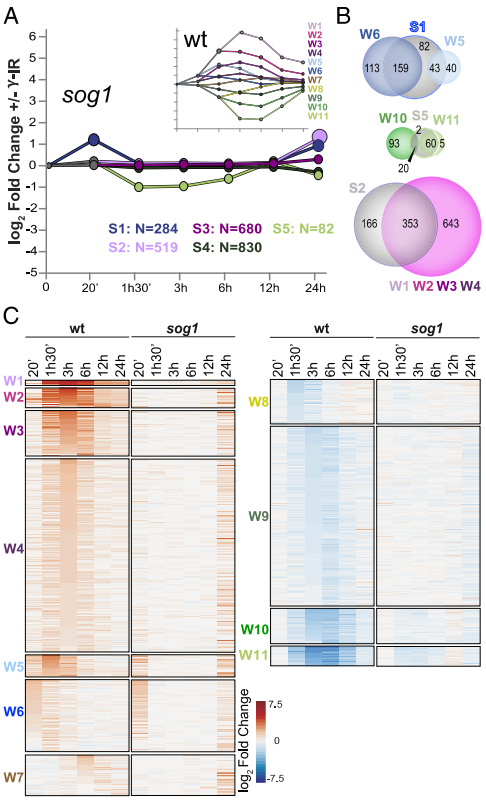


Fig. 2. SOG1 controls nearly all aspects of the transcriptional response to γ-IR. (*A*) *sog1* DREM model [see Source Data 1 (44)] showing five sets of coexpressed genes, termed *sog1* paths S1–S5. The expression profiles, enriched GO terms, and motifs are presented in *SI Appendix*, Figs. S6 and S9, respectively. All genes are listed in Dataset S3*B*. For comparison, the wild-type (wt) DREM model is shown as an inlay. (*B*) Scaled Venn diagrams showing the overlap of genes in DREM paths with similar trends in the wild-type and *sog1* models. (*C*) Heatmaps showing the $\log_2$ FC in expression (γ-IR vs. mock) of the genes present in paths W1–W11 (Fig. 1*A*) using either the wild-type or the *sog1* expression data. For each path, the heatmaps were ranked based on the wild-type expression level. See also *SI Appendix*, Fig. S7.

6 h (Figs. 1*A* and 2*C*). However, essentially none of these genes are induced at the proper time (between 1 h 30 min and 6 h) in the *sog1* mutant (Fig. 2*C*). As with the other features revealed by the *sog1* DREM model, these expression profiles were independently verified (*SI Appendix*, Fig. S7). While many of the late-induced genes in the *sog1* mutant correspond to general stress-response genes, several well-known DNA repair genes, including *GMI1*, *BRCA1*, *PARP2*, and *PARG2* are also present [Dataset S3*B* and Source Data 2 (44)]. Thus, we posit that this latent response may be triggered by defects in DNA repair and the persistence of unrepaired DNA lesions that may be accentuated by the loss of SOG1 function.

In sum, our analysis of the DNA damage response in *sog1* mutants demonstrate that, with the exception of an initial burst in the expression of general stress-response genes, essentially all other aspects of the wild-type DNA damage response are either fully or partially dependent on SOG1. This includes the 1,233 genes present in paths W1–W5 and W7 that require SOG1 for their induction between 20 min and 12 h, the 712 genes present in paths W8 and W9 that require SOG1 for their repression, and the 178 genes present in paths W10 and W11 that show a partial dependence on SOG1. These findings greatly expand the set of genes known to be induced in a SOG1-dependent manner and demonstrate that SOG1 is a master regulator not only for genes induced by DNA damage, but also for those that are repressed.

SOG1 Is a Transcriptional Activator that Directly Targets Nearly Half of the Genes Strongly Induced by DNA Damage. As described above, SOG1 is required for the induction or repression of thousands of genes in response to γ-IR. Until recently (27), only a few SOG1 target genes had been identified, namely *SIAMESE-RELATED 5* (*SMR5*) and *SIAMESE-RELATED 7* (*SMR7*) (26), *FLAVIN-DEPENDENT MONOOXYGENASE 1* (*FMO1*) (22), *CYCLIN B1;1* (*CYCB1;1*) (28), and *BREAST CANCER SUSCEPTIBILITY 1* (*BRCA1*) (25). To determine which of the other genes regulated by SOG1 are direct targets, chromatin immunoprecipitation and sequencing (ChIP-seq) experiments were conducted (Dataset S4*A*) using transgenic lines in which the *sog1* mutation was complemented by a SOG1-3xFLAG construct driven by the endogenous SOG1 promoter (*SI Appendix*, Fig. S10 *A* and *B*). As SOG1 binding is expected to precede the transcriptional regulation of its targets, the ChIP-seq experiments were performed at two early time points after irradiation, 20 min and 1 h. Analysis of the resulting ChIP-seq profiles identified 307 SOG1 peaks (Dataset S4*B*) that are mainly located in promoters and transcribed regions (*SI Appendix*, Fig. S10*C*) and were assigned to 310 immediately adjacent gene targets (Dataset S4 *C* and *D*). These peaks were enriched relative to both input and wild-type ChIP samples (*SI Appendix*, Fig. S10*D*) and, attesting the reproducibility and quality of the ChIP-seq experiments, similar enrichment patterns were observed for both the 20-min and 1-h datasets (Fig. 3*A*). Furthermore, four of the five initially identified SOG1 targets (*SI Appendix*, Fig. S10*E*), and

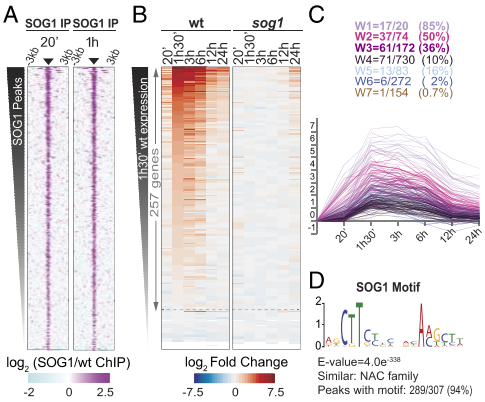


Fig. 3. SOG1 is a transcriptional activator that directly regulates nearly half of the genes strongly induced by γ-IR. (*A*) Heatmap showing SOG1 enrichment [$\log_2$(SOG1/wt ChIP)] at the 307 peaks (±3 Kb) identified from both ChIP assays (20 min and 1 h) (*SI Appendix*, Fig. S10 and Dataset S4*B*). (*B*) Heatmaps showing the expression of SOG1 target genes (Datasets S4 *C* and *D*) during the wild-type (wt) and *sog1* γ-IR time courses ranked based on the wild-type expression values at 1 h 30 min. (*C*) Expression profiles of the SOG1 target genes colored by gene paths. The fraction of genes in each path is indicated, with the percentage in parentheses. (*D*) Motif identified under the SOG1 ChIP peaks. The E-value, as reported by MEME (97, 102), and the fraction of peaks with the motif are indicated below (see also Dataset S4*E*).

PLANT BIOLOGY

104 of the 146 recently identified SOG1 ChIP-seq targets (27), were among the 310 SOG1 targets identified after γ-IR treatment (*SI Appendix*, Fig. S10*F*). Thus, the SOG1 ChIP-seq results presented here are consistent with, but greatly expand upon, the previously identified SOG1 targets.

To determine whether genes targeted by SOG1 are induced, repressed, or both in response to DNA damage, their expression patterns after γ-IR treatment were determined from the wild-type and *sog1* γ-IR time-course datasets (Fig. 3*B*). Over 80% of these genes were induced by γ-IR in a SOG1-dependent manner, while none were significantly repressed (Fig. 3*B*), demonstrating that SOG1 acts, either by itself or in association with other factors, as a transcriptional activator. In agreement with this finding, 206 of the 310 target genes were identified as part of the wild-type DREM model. Not only do all these genes map to up-regulated gene paths (Fig. 3*C* and *SI Appendix*, Fig. S10*G*), they correspond to 43% of all of the strongly up-regulated genes (e.g., W1–W3) and 85% of the most highly induced genes (e.g., W1) (Fig. 3*C*). Strikingly, the same motif identified from the path W1–W3 genes of the DREM network (CTT[N_7]AAG) (Fig. 1*E*) was found directly under nearly all (289 of 307) of the SOG1 peaks (Fig. 3*D* and Dataset S4*E*). This motif is similar to motifs observed for other NAC TFs (48, 51, 52) and is also consistent with the findings of Ogita et al. (27), as this motif was identified in the promoter regions of approximately half of their SOG1 target genes and was shown to be necessary and sufficient for SOG1 binding based on in vitro and in vivo assays. Together, these analyses reveal several hundred direct targets of SOG1 and demonstrate that although SOG1 is required for both the induction and repression of DNA damage-responsive genes, it acts strictly as a transcriptional activator, directly controlling ∼8% of the DNA damage transcriptional response.

Further investigation into the direct targets of SOG1 demonstrated that it regulates genes involved in many DNA damage-associated processes, including transcriptional and posttranscriptional regulation, oxidative stress, defense, cell cycle regulation, cell death, and DNA repair (Fig. 4). Among these categories, the most strongly induced genes correspond to DNA repair proteins, cell cycle regulators, and TFs (Fig. 4). For DNA repair, many loci besides the initially identified repair-associated genes, *BRCA1* (25) and *CYCB1;1* (28), or even the more recently identified genes (27), were identified as SOG1 targets (Fig. 4). These targets encode proteins that are required for homologous recombination (HR) in plants and have been assigned to different repair steps in yeast, mammals, and/or plants, including early signaling events downstream of the MRN complex (*RAD17*) (58), DNA end resection (*GR1*/*COM1*) (59), interhomolog pairing [*HOP2* (60), *MND1* (61, 62), and *XRI1* (63)], strand invasion (*RAD51* and *RAD54*) (9), and holiday junction formation/resolution [*RFC1* (64, 65) and *RecQl3* (66)]. In addition, SOG1 targets factors that affect HR, but play roles that are less well understood, including *NSE4A* (67), *AGO2* (68), and *GMI1* (32). Besides HR factors, SOG1 targets DNA polymerases involved in DNA replication (*POLD4*) (69) and DNA repair [*POLK* (70) and *POL2A* (71)], as well as genes involved in the production of deoxyribonucleotides [*TSO2*, *RNR1* (72) and *TK1a* (73, 74)] that have been implicated in DNA repair (72, 75). Finally, several SOG1 targets remain uncharacterized, but are homologous to repair factors in other organisms, or contain domains indicating that they might have a role in DNA repair, including *AT1G80850*, a gene encoding a DNA glycosylase (76), *AT4G02110*, a gene encoding a BRCA1 C-terminal domain protein (77), and *FANCJA*, a gene homologous to a mammalian Fanconi anemia group I-like factor involved in HR repair (78) (Fig. 4). Together, these analyses demonstrate that SOG1 rapidly (within 20 min) and directly targets key HR and DNA metabolism factors to facilitate DNA repair in a coordinated fashion.

Consistent with SOG1 only directly targeting ∼8% of the DREM network, there are 33 TF genes, from 14 different TF families, that are directly regulated by SOG1, revealing the first tier of a complex cascade of transcriptional responses initiated by SOG1 (Fig. 4 and *SI Appendix*, Fig. S11). Among these SOG1 targets are six other NAC TFs that are highly induced by γ-IR, highlighting the prominent role of this TF family in the DNA damage response. SOG1 also directly

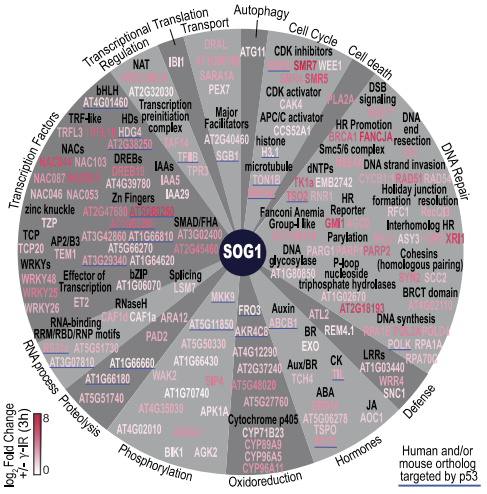


Fig. 4. Functional categorization of SOG1 target genes. Of the 206 SOG1 target genes in the DREM model, 141 were assigned to functional categories and colored based on the $\log_2$ FC +/− γ-IR at the 3-h time point from the wild-type DREM model. Genes underlined in blue have a human and/or mouse ortholog shown to be targeted and up-regulated by p53. See also *Materials and Methods*, *SI Appendix*, Figs. S11 and S12, and Source Data 4 (44).

regulates *TCP20*, which encodes a TF that binds the *CYCB1;1* promoter (79) and may help reinforce its up-regulation. Further extrapolation of the SOG1 TF network based on previously annotated gene–TF interactions from the AGRIS (45–47) and DAP-seq databases (48) allows the inference of a putative second tier within the network (*SI Appendix*, Fig. S11). Although this second tier remains to be experimentally validated, it provides a framework for the generation and examination of additional hypotheses regarding the transcriptional response to DNA damage.

In agreement with SOG1 being required for cell cycle arrest in response to DNA damage (12), four cyclin-dependent kinase inhibitor (CKI) genes, belonging to two classes, [KIP-RELATED PROTEINS (KRP) and SIAMESE/SIAMESE-RELATED (SMR)] (80, 81), were identified as SOG1 targets here (Fig. 4) and in Ogita et al. (27). Current models posit that the up-regulation of CKIs by SOG1 plays an important role in stabilizing the Rep-MYB3R family of TFs, which are required for the repression of G2/M-regulated genes in response to DNA damage (53). As such, these SOG1 ChIP-seq findings expand the pool of CKIs that may be acting upstream of the Rep-MYB3Rs to include not just the initially identified *SMR5* and *SMR7* targets (26), but also *SMR4* and *KRP6* (*SI Appendix*, Fig. S12). Interestingly, SOG1 was also found to directly target *WEE1*, which encodes a kinase that is required for proper cell cycle arrest in response to replicative stress (82), but not in response to DSBs (15), suggesting that SOG1 activates cell cycle arrest programs associated with different types of damage. Finally, in addition to cell cycle arrest, SOG1 is also required for the promotion of endoreduplication in response to DNA damage (15). Although SOG1 does not directly regulate genes with known roles in the endocycle, it does target and induce the expression of *WRKY25*. Because this TF was shown by DAP-seq to bind the promoter of *DP-E2F-LIKE 1* (*DEL1*) (48), a gene important for the repression of endoreduplication (83) (*SI Appendix*, Fig. S12), it can be hypothesized that WRKY25, or other related WRKY TFs, represses *DEL1* expression in response to DNA damage to promote endoreduplication. Based on these findings it is clear that,

although much of the molecular details remain to be elucidated, SOG1 plays critical, but indirect roles in controlling the suppression of cell cycle genes, which is consistent with the DREM model where the induction of DNA repair factors and other direct targets of SOG1 precede the suppression of cell cycle genes.

Finally, in addition to previously drawn parallels between SOG1 and the mammalian p53 protein, which focused on the activation of SOG1 by ATM and the common DNA damage-associated processes dependent on these two TFs (cell cycle arrest, cell death, overall genome stability, and the induction of damage-response genes), the identification and analysis of SOG1 target genes has revealed additional parallels. First, both proteins act as transcriptional activators (84, 85). Second, they target genes related to similar biological processes (19). And third, many of the SOG1 target genes have human and/or mouse orthologs identified as p53 targets (Fig. 4), including the RNR subunit, *TSO2*, for dNTP balance maintenance (86); the DNA polymerase kappa, *POLK*, for translesion DNA synthesis (87); the histone variant *H3.1*, which is deposited in a DNA-synthesis–dependent manner and is incorporated at damaged chromatin (88); and *KRP6*, which contains a cyclin-dependent kinase inhibitor domain similar to that of *p21*, a mammalian gene that mediates the p53-dependent down-regulation of cell cycle genes (89). However, SOG1 is unique in its selective targeting of numerous genes required for repair by HR (Fig. 4) (27). Thus, despite the fact that there is no sequence conservation between p53 and SOG1, they share a subset of conserved target genes, suggesting that they have been coopted to mediate both shared and unique aspects of the DNA damage response in plants versus mammals.

The Rep-MYB3R Family Is Required to Suppress Cell Cycle Genes after DNA Damage. Although the direct targets of SOG1 are activated in response to DNA damage, hundreds of repressed genes also depend on SOG1. Thus, events set into motion by the expression of SOG1 targets must play key roles in repressing genes in response to DNA damage. Of particular interest are those that regulate cell cycle genes, which are strongly repressed in the wild-type DREM model (e.g., W10 and W11). Consistent with the TF predictions from our DREM model (Fig. 1*A* and *SI Appendix*, Fig. S4), recent studies have revealed connections between the Rep-MYB3R family, cell cycle regulation, and DNA damage. First, all three family members—MYB3R1, MYB3R3, and MYB3R5—were found to act redundantly to suppress the expression of 34 genes associated with the G2/M phase of the cell cycle (90), 29 of which fall into paths W9, W10, and W11 (*SI Appendix*, Fig. S13*A*). Second, the *myb3r3* and *myb3r5* mutants display enhanced tolerance to DNA damage agents, including γ-IR, and show defects in cell cycle regulation and cell death (53). Finally, although the Rep-MYB3Rs are not transcriptionally regulated in response to DNA damage (*SI Appendix*, Fig. S13*B*) (53), they were placed in a SOG1-dependent pathway based on epistasis experiments (53). Together, these findings demonstrate that the Rep-MYB3Rs are essential for inhibiting cell division during the DNA damage response in connection with SOG1, yet only a few genes repressed in a MYB3R1/3/5-dependent manner after DNA damage have been identified (53).

To identify genes regulated by the Rep-MYB3R family in response to DNA damage, mRNA-seq experiments were conducted in wild-type and *myb3r1,3,5* triple mutants 3 h after either mock or γ-IR treatments (*SI Appendix*, Fig. S13*B* and Dataset S1), a time when hundreds of genes are strongly down-regulated in the wild-type DREM model (Fig. 1*A*). In agreement with previous studies showing minimal expression changes between wild-type and *myb3r1,3,5* triple mutants in early seedling stages (90), comparison of the 6-d-old mock-treated seedlings (wild-type vs. *myb3r1,3,5*) revealed only 24 up-regulated genes, including just two G2/M phase genes and a single down-regulated gene (Dataset S5*A*). However, after γ-IR treatment, the DNA damage response was clearly altered in the *myb3r1,3,5* triple mutant compared with the wild-type control. On a global level, the γ-IR response observed in the wild-type dataset was similar to the 3-h time point from the wild-type DREM model, but in the *myb3r1,3,5* triple dataset, the genes in two of the three cell cycle-enriched paths (W10 and W11) were less repressed overall (Fig. 5*A*). At the level of individual genes, 80 loci significantly less repressed in the *myb3r1,3,5* mutants after DNA damage (Dataset S5 *B* and *C*) (FC ≥ 2 and FDR ≤ 0.05) were determined by considering both the experimental conditions (γ-IR vs. mock treatments) and the genotypes (wild-type vs. *myb3r1,3,5*). Nearly all of these genes (78 of 80) are present in the wild-type DREM model, constituting 72.3% of the path W11 genes (47 of 65), 24.8% of the path W10 genes (28 of 113), and 0.5% of the path W9 genes (3 of 571) (Fig. 5*B* and *SI Appendix*, Fig. S13*C*).

Functionally, 71 of these 80 genes are associated with the G2/M phase of the cell cycle (54, 57) (Fig. 5*C* and Dataset S5*B*). Approximately one-third of these genes (28 of 70) were previously shown to be repressed in a Rep-MYB3R–dependent manner either under normal growth conditions (90) or after exposure to DNA damage (53) (Dataset S5*B*). However, the remaining two-thirds (42 of 70) represent newly identified Rep-MYB3R–regulated genes (Dataset S5*B*). Finally, these 80 genes are likely direct targets of the Rep-MYB family, as they almost all (72 of 80) possess MSA motifs in their promoters and/or are associated with previously defined MYB3R3 peaks by ChIP-seq (q-value > 25 under nondamaged conditions) (90) or by ChIP-qPCR after DNA damage (53) (Fig. 5*D* and *SI Appendix*, Fig. S13*D*). Furthermore, the association of MYB3R3 with these

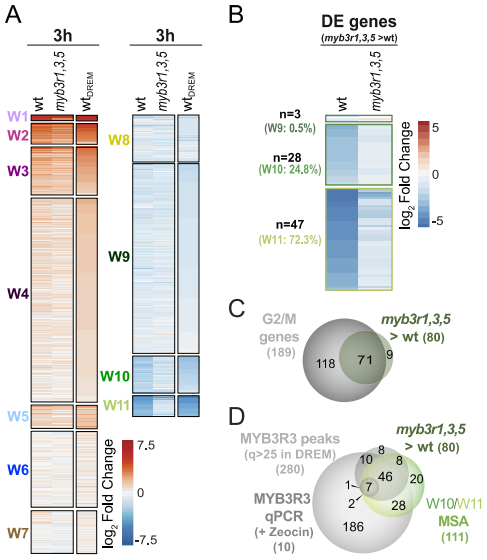


Fig. 5. The Rep-MYB3R TFs are the master repressors of cell cycle genes in response to γ-IR. (*A*) Heatmaps showing the $\log_2$ FC in expression (γ-IR vs. mock) of the genes present in paths W1–W11 of the wild-type (wt) DREM model, ordered as in Fig. 2, using either the wild-type or the *myb3r1,3,5* expression data. For reference, the expression levels from the wild-type DREM model (wt_{DREM}) at the 3-h time point was included. (*B*) Heatmaps showing the $\log_2$ FC in expression (γ-IR vs. mock) of the path W9, W10, and W11 genes from the wild-type DREM model that are significantly less repressed in the *myb3r1,3,5* mutant than in the wild-type controls ("*myb3r1,3,5* > wt") (Dataset S5 *B* and *C*). (*C* and *D*) Scaled Venn diagrams showing the overlap between the genes shown in *B* and either genes expressed in the G2/M phase of the cell cycle (54, 57) (*C*) or genes associated with MSA motifs and/or MYB3R3 peaks (53, 90) (*D*).

PLANT BIOLOGY

genes is quite specific, as the same DREM paths affected in the *myb3r1,3,5* triple mutants displayed the highest levels of promoter-proximal MYB3R3 enrichment (*SI Appendix*, Fig. S13*E*). In contrast to the genes less repressed in the *myb3r1,3,5* mutants after DNA damage, the genes significantly less induced showed poor overlap with genes in the wild-type DREM model (*SI Appendix*, Fig. S13*C*) and none overlapped with G2/M-expressed genes or MYB3R3 peaks (Dataset S5*B*), suggesting that these genes may not be directly regulated by the Rep-MYB3R family. Taken together, these findings more than double the number of cell cycle–regulated genes known to be regulated by the Rep-MYB3R family and demonstrate that these TFs control a large portion of the most strongly repressed genes in the wild-type DREM network in response to DNA damage.

Conclusions

Despite the identification of TFs that control most of the cellular outcomes of the DNA damage response, namely p53 in mammals and SOG1 in plants (8), knowledge regarding the gene-expression networks controlled by these TFs, and the dynamics of the DNA damage response as a whole, has remained poorly understood. Here, DREM analysis of the transcriptional changes occurring in response to γ-IR resulted in the classification of the *Arabidopsis* DNA damage response into 11 groups of coexpressed genes with distinct expression profiles, biological functions, and *cis*-regulatory features. Attesting to the power of this approach, this network captured the major biological processes associated with the DNA damage response, including DNA metabolism, DNA repair, cell cycle regulation, and cell death, while at the same time providing new insight into early responses that occur in a largely SOG1-independent manner and late responses that occur specifically in the absence of SOG1. In addition, the DREM model revealed the order of key transcriptional events, including an initial burst in the expression of general stress genes, a rapid and sustained induction of DNA repair and DNA metabolism genes, and a slightly delayed suppression of cell cycle-associated genes. Finally, analysis of additional genetic and genomic experiments within the context of this model enabled the assignment of SOG1 and the Rep-MYB3R TFs to the most strongly up- and down-regulated gene clusters within the network, demonstrating that they act as the major activator and repressors of the DNA damage response, respectively (*SI Appendix*, Fig. S14).

The gene-expression profiles and TFs identified based on this DREM analysis represent a transcriptional roadmap of the DSB response in *Arabidopsis* that serves as a framework to continue exploring the functions of both known and unknown genes. Such studies may prove especially useful for genes present in paths with highly enriched GO terms, as uncharacterized genes coexpressed with, for example, DNA repair factors or cell cycle regulators, may have similar functions. Similarly, this map will facilitate the identification and characterization of additional TFs and expression programs acting downstream of, or independent from, SOG1 to coordinate the diverse processes associated with the DNA damage response. In moving forward, it will also be important to continue integrating information into the DREM model for factors that, like the Rep-MYB3Rs and SOG1, are regulated at the posttranscriptional rather than the transcriptional level (16, 17, 53). Indeed, analysis of the *Arabidopsis* phospho-proteome in response to ionizing radiation identified hundreds of proteins phosphorylated upon DNA damage, including many TFs (91), but the functions of most of these posttranslational modifications remain unknown. Finally, it will be interesting to extend upon previous comparisons between the transcriptional responses elicited by various genotoxic stresses (41), via similar network approaches with matched experimental designs, to gain a higher-resolution view of how the profiles differ in response to specific types of DNA damage. Continued efforts toward addressing the aforementioned aspects of the DNA damage response will shed additional light on the inner workings of this process, which can ultimately be leveraged for crop improvement, because many environmental factors—including UV-B exposure, toxic levels of aluminum or boron in the soil, or endogenous processes like seed aging and germination—are associated with various types of DNA damage and show genetic requirements and transcriptional responses with links to the DNA damage response induced by DSBs (5, 41, 92).

Materials and Methods

Plant Materials. *Arabidopsis* lines include the Columbia (Col) wild-type strain, the *sog1-1* EMS mutant provided by A. Britt, Department of Plant Biology, University of California, Davis, CA, which is in a mixed Landsberg *erecta* (Ler)/Col background containing a *CYCB1;1 promoter::GUS* fusion (13), and the *myb3r1,3,5* triple mutant provided by M. Ito, Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya, Japan (90).

mRNA-seq Experiments.

Growth conditions and library preparation. For all mRNA-seq experiments, seeds were sterilized, sown on half MS media with 0.6% plant agar, stratified at 4 °C for 3 d, and transferred into a growth chamber [constant light (100 μmol·m^{-2}·s^{-1}) and temperature (23 °C)]. For the wild-type, *sog1*, and *myb3r1,3,5* datasets, 6-d-old seedlings were γ-IR at a dose of 100 Gy using a Co60 radioactive source. See *SI Appendix, Materials and Methods* for details regarding dose rates and experimental designs. Total RNA was extracted from pooled seedlings using the Quick-RNA MiniPrep kit (Zymo Research R1055) and mRNAs were purified from 1 to 2 μg of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs E7490) before library preparation with the NEBNext Ultra or UltraII RNA Library Prep Kit (New England Biolabs E7530 or E7770). All sequencing was performed on an Illumina HiSEq 2500 in 50-bp single-end mode.

Data processing. For all mRNA-seq experiments, Illumina reads were mapped to the TAIR10 genome using STAR (93) (Dataset S1). Expression values for all genes and samples were retrieved using the analyzeRepeat.pl script from HOMER (94) with the -noadj option. These values were used to generate the principle component analysis plots shown in *SI Appendix*, Fig. S1*A* with ggplot in R studio and to identify DE genes (FC ≥ 2 with an FDR ≤ 0.01) based on the experimental conditions (i.e., γ-IR vs. mock) (Dataset S2) or genotype (i.e., wild-type vs. mutant) (Dataset S5*A*) using the getDiffExpression.pl script from HOMER (94) with the -DESeq2 and -repeats options. To identify DE genes based on both the experimental conditions and genotypes for the "wt vs *myb3r1,3,5*" experiment (Dataset S5 *B* and *C*), the analysis was performed in R studio using DESeq2 with design = batch + Condition + Group + Condition:Group. To visualize the gene-expression data, University of California, Santa Cruz (UCSC) browser tracks were generated using the makeUCSCfile script from HOMER (94), normalizing to 10 million reads. Tracks showing the difference between the average mock- and γ-IR–treated samples (Fig. 1*B*) were generated using deepTools (95) as detailed in *SI Appendix, Materials and Methods*. Heatmaps showing the expression of individual genes during the γ-IR time course were generated using the normalized $\log_2$ FC values generated by DESeq2 and plotted in R studio using pheatmap. Venn diagrams comparing DE gene sets were generated using VennMaster (96).

DREM Analysis. For data inputs and models, see Source Data 1 (44). Three types of files were generated as inputs for the DREM models. The first is a list of the 33,323 TAIR10 gene IDs, the second is a set of gene–TF interactions, and the third contains the expression values from the RNA-seq experiments. See *SI Appendix, Materials and Methods* for details regarding the content and generation of these files. As detailed in *SI Appendix, Materials and Methods*, the input files described above were used to generate the wild-type and *sog1* DREM models by a two-step process using the DREM v2.0.3 (42, 43).

Motif Analysis. As detailed in *SI Appendix, Materials and Methods*, the MEME suite (97) was used to identify motifs enriched in the promoters of genes present in the individual DREM paths. All found motifs were then compared with the DAP-seq database of motifs (48) using Tomtom (98) to search for motif matches overlapping by at least 5 bp with a distance coefficient calculated by the Pearson method showing an E-value < 10. Only the matches with an E-value < 0.01 were kept.

GO Analysis. As detailed in *SI Appendix, Materials and Methods*, the GO enrichments for each DREM path were retrieved using GO-TermFinder (99) via the Princeton GO-TermFinder interface (https://go.princeton.edu/cgi-bin/GOTermFinder), simplified using the REVIGO (100) (revigo.irb.hr/), and visualized as unclustered heatmaps in Treeview.

SOG1 ChIP-Seq Experiments.

ChIP. For ChIP, 1.3 g of 10-d-old seedlings from wild-type and SOG1-3xFLAG transgenic lines were harvested after 20 min and 1 h of γ-IR (75 Gy at a dose-rate of 3.75 Gy/min), ground in liquid nitrogen, and resuspended in 25 mL of nuclear isolation buffer. Proteins were cross-linked first with 1.5 mM ethylene glycol bis(succinimidyl succinate) (Thermo Fisher Scientific 21565) for 20 min at room temperature, and then with formaldehyde at a final concentration of 1% for 10 min at room temperature. Cross-linking was stopped by adding 1.7 mL of 2 M glycine and incubating for 10 min at room temperature. The cross-linked samples were filtered through Miracloth and centrifuged at 4,500 × *g* for 20 min at 4 °C. Pellets were resuspended in 1-mL extraction buffer 2 and centrifuged at 12,500 × *g* for 10 min at 4 °C. Pellets were resuspended in 300 μL of nuclei lysis buffer. The chromatin was sonicated using a Bioruptor (Diagenode) for 20 cycles (30-s on/30-s off, on high intensity) at 4 °C. The debris was pelleted by centrifugation at 9,600 × *g* for 10 min at room temperature; 30 μL were kept as an input. The remaining sample was diluted 10 times using ChIP dilution buffer. Next, 25 μL of M2-FLAG magnetic bead slurry (Sigma M8823) was added and the sample was rotated for 1 h at 4 °C. The beads were washed sequentially in 1 mL of low-salt buffer, 1 mL of high-salt buffer, 1 mL LiCl buffer, and 1 mL of TE. The chromatin elution was performed in 300 μL SDS elution buffer at 65 °C and the cross-linking was reversed by adding 12 μL of 5 M NaCl and 30 μL of 1 M DTT and incubating the samples overnight at 65 °C. Digestion with Proteinase K was performed at 45 °C for 1 h. The DNA was purified by phenol-chloroform-IAA extraction and precipitated with ethanol. Pellets of both inputs and IPs were resuspended in 30 μL of TE, pH 8.0. All buffers are described in Dataset S6.

Library preparation, sequencing, and analysis. Libraries were prepared from 25 μL of the IP samples or 1 μg of the input samples using the NEBNext Ultra II DNA Library Prep Kit (New Englabnd Biolabs E7645). Sequencing was performed on an Illumina HiSEq 2500 in 50-bp single-end mode. Reads were mapped to the TAIR10 genome using bowtie2 (101) (Dataset S4A), UCSC browser tracks normalized to 10 million reads were generated using HOMER (94), and, as detailed in *SI Appendix, Materials and Methods*, heatmaps and metaplots showing SOG1 ChIP enrichments were generated using the deepTools suite (95). SOG1 peaks at 20 min or 1 h were called independently relative to two controls (the wt_IP and input samples) using the HOMER findPeaks tool (94) and only peaks identified relative to both controls were kept. For details regarding peak calling and gene assignments, see *SI Appendix, Materials and Methods*. Heatmaps showing the expression of SOG1 target genes (Fig. 3*B*) were generated using the $\log_2$ FC values generated by DESeq2 and plotted in R studio using pheatmap and the overlaps between SOG1 target genes identified here or in ref. 27 were generated using VennMaster (96). Peak locations relative to genome features (promoters, introns, exons, and so forth) were determined using the HOMER annotatePeaks.pl script (94), where the promoter regions are defined as −1 kb and +100 bp relative to the transcription start site. Motifs under the SOG1 peaks were determined using the MEME tool (102) from the MEME suite (97) with the following options (-nostatus -maxsize 7500000 -nmotifs 10 -minw 6 -maxw 18 -revcomp -psp -bfile).

MYB3R3 ChIP-Seq Analysis. Using data from GSE60554 (90), the closest TAIR10 gene to each peak was determined using annotatepeaks.pl from HOMER (94). Mapping of the ChIP-seq reads and analysis of ChIP enrichment profiles using deepTools were as described for the SOG1 ChIP-seq. Venn diagrams were generated using VennMaster (96) based on known G2/M-expressed genes (54, 57) or previously defined MYB3R3 ChIP peaks (90).

Cytoscape. Using Cytoscape 3.4.0, networks for either the 141 SOG1 target genes that were assigned to functional categories based on the GO analysis and their TAIR10 annotations [Source Data 4 (44)] or the DREM network targets of the 33 TFs downstream of SOG1 identified from the AGRIS (45–47) and DAP-seq. (48), databases [*SI Appendix*, Fig. S11 and Source Data 4 (44)] were constructed and colored based on the $\log_2$ FC in expression 3 h after γ-IR. For Fig. 4, genes underlined in blue represent those that have a human or mouse ortholog, identified using the PANTHER (103) and Thalemine tools (104), that were shown to be targeted and up-regulated by p53 based on 13 genome-wide studies in humans (105) or a single study in primary mouse embryo fibroblasts study (106).

ACKNOWLEDGMENTS. We thank the J.A.L. laboratory and colleagues at the Salk Institute for helpful discussions; and Dr. C. Huang, Dr. L. Song, and Dr. Y. He for bioinformatics support. Work in the J.A.L. laboratory was supported by the Rita Allen and Hearst Foundations (to J.A.L.). C.B. was supported by the Catharina Foundation Fellowship. N.V. was supported by the Jesse and Caryl Philips Foundation. This work was also supported by the Next-Generation Sequencing Core Facility and the Integrative Genomics and Bioinformatics Core Facility at the Salk Institute, with funding from NIH-National Cancer Institute Cancer Center Support Grant P30 014195, the Glenn Foundation for Medical Research, the Chapman Foundation, and the Helmsley Charitable Trust.

1. Ciccia A, Elledge SJ (2010) The DNA damage response: Making it safe to play with knives. *Mol Cell* 40:179–204.
2. Manova V, Gruszka D (2015) DNA damage and repair in plants—From models to crops. *Front Plant Sci* 6:885.
3. Aguilera A, García-Muse T (2013) Causes of genome instability. *Annu Rev Genet* 47:1–32.
4. Rodgers K, McVey M (2016) Error-prone repair of DNA double-strand breaks. *J Cell Physiol* 231:15–24.
5. Hu Z, Cools T, De Veylder L (2016) Mechanisms used by plants to cope with DNA damage. *Annu Rev Plant Biol* 67:439–462.
6. Su TT (2006) Cellular responses to DNA damage: One signal, multiple choices. *Annu Rev Genet* 40:187–208.
7. Amiard S, Gallego ME, White CI (2013) Signaling of double strand breaks and deprotected telomeres in *Arabidopsis*. *Front Plant Sci* 4:405.
8. Yoshiyama KO, Sakaguchi K, Kimura S (2013) DNA damage response in plants: Conserved and variable response compared to animals. *Biology (Basel)* 2:1338–1356.
9. Bleuyard JY, Gallego ME, White CI (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair (Amst)* 5:1–12.
10. Shiloh Y, Ziv Y (2013) The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* 14:197–210.
11. Knoll A, Fauser F, Puchta H (2014) DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosome Res* 22:191–201.
12. Preuss SB, Britt AB (2003) A DNA-damage-induced cell cycle checkpoint in *Arabidopsis*. *Genetics* 164:323–334.
13. Yoshiyama K, Conklin PA, Huefner ND, Britt AB (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci USA* 106:12843–12848.
14. Furukawa T, et al. (2010) A shared DNA-damage-response pathway for induction of stem-cell death by UVB and by gamma irradiation. *DNA Repair (Amst)* 9:940–948.
15. Adachi S, et al. (2011) Programmed induction of endoreduplication by DNA double-strand breaks in *Arabidopsis*. *Proc Natl Acad Sci USA* 108:10004–10009.
16. Yoshiyama KO, et al. (2013) ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in *Arabidopsis*. *EMBO Rep* 14:817–822.
17. Yoshiyama KO, Kaminoyama K, Sakamoto T, Kimura S (2017) Increased phosphorylation of Ser-Gln sites on SUPPRESSOR OF GAMMA RESPONSE1 strengthens the DNA damage response in *Arabidopsis thaliana*. *Plant Cell* 29:3255–3268.
18. Yoshiyama KO (2016) SOG1: A master regulator of the DNA damage response in plants. *Genes Genet Syst* 90:209–216.
19. Kastenhuber ER, Lowe SW (2017) Putting p53 in context. *Cell* 170:1062–1078.
20. Meek DW, Anderson CW (2009) Posttranslational modification of p53: Cooperative integrators of function. *Cold Spring Harb Perspect Biol* 1:a000950.
21. Biever JJ, Brinkman D, Gardner G (2014) UV-B inhibition of hypocotyl growth in etiolated *Arabidopsis thaliana* seedlings is a consequence of cell cycle arrest initiated by photodimer accumulation. *J Exp Bot* 65:2949–2961.
22. Chen P, Umeda M (2015) DNA double-strand breaks induce the expression of flavin-containing monooxygenase and reduce root meristem size in *Arabidopsis thaliana*. *Genes Cells* 20:636–646.
23. Hu Z, Cools T, Kalhorzadeh P, Heyman J, De Veylder L (2015) Deficiency of the *Arabidopsis* helicase RTEL1 triggers a SOG1-dependent replication checkpoint in response to DNA cross-links. *Plant Cell* 27:149–161.
24. Pedroza-García JA, et al. (2017) Function of the plant DNA polymerase epsilon in replicative stress sensing, a genetic analysis. *Plant Physiol* 173:1735–1749.
25. Sjogren CA, Bolaris SC, Larsen PB (2015) Aluminum-dependent terminal differentiation of the *Arabidopsis* root tip is mediated through an ATR-, ALT2-, and SOG1-regulated transcriptional response. *Plant Cell* 27:2501–2515.
26. Yi D, et al. (2014) The Arabidopsis SIAMESE-RELATED cyclin-dependent kinase inhibitors SMR5 and SMR7 regulate the DNA damage checkpoint in response to reactive oxygen species. *Plant Cell* 26:296–309.
27. Ogita N, et al. (2018) Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in *Arabidopsis*. *Plant J* 94:439–453.
28. Weimer AK, et al. (2016) The plant-specific CDKB1-CYCB1 complex mediates homologous recombination repair in *Arabidopsis*. *EMBO J* 35:2068–2086.
29. Charbonnel C, Gallego ME, White CI (2010) Xrcc1-dependent and Ku-dependent DNA double-strand break repair kinetics in *Arabidopsis* plants. *Plant J* 64:280–290.
30. Charbonnel C, Allain E, Gallego ME, White CI (2011) Kinetic analysis of DNA double-strand break repair pathways in *Arabidopsis*. *DNA Repair (Amst)* 10:611–619.
31. Friesner JD, Liu B, Culligan K, Britt AB (2005) Ionizing radiation-dependent gamma-H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. *Mol Biol Cell* 16:2566–2576.
32. Böhmdorfer G, et al. (2011) GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in *Arabidopsis*. *Plant J* 67:420–433.
33. Culligan KM, Robertson CE, Foreman J, Doerner P, Britt AB (2006) ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J* 48:947–961.
34. Gicquel M, Taconnat L, Renou JP, Esnault MA, Cabello-Hurtado F (2012) Kinetic transcriptomic approach revealed metabolic pathways and genotoxic-related changes implied in the *Arabidopsis* response to ionising radiations. *Plant Sci* 195:106–119.

35. Kim JH, et al. (2007) Transcriptomic profile of *Arabidopsis* rosette leaves during the reproductive stage after exposure to ionizing radiation. *Radiat Res* 168:267–280.
36. Kim JB, et al. (2014) Differentially expressed genes in response to gamma-irradiation during the vegetative stage in *Arabidopsis thaliana*. *Mol Biol Rep* 41:2229–2241.
37. Nagata T, Yamada H, Du Z, Todoriki S, Kikuchi S (2005) Microarray analysis of genes that respond to gamma-irradiation in *Arabidopsis*. *J Agric Food Chem* 53:1022–1030.
38. Wang Z, Schwacke R, Kunze R (2016) DNA damage-induced transcription of transposable elements and long non-coding RNAs in *Arabidopsis* is rare and ATM-dependent. *Mol Plant* 9:1142–1155.
39. Kim DS, et al. (2011) Antioxidant response of *Arabidopsis* plants to gamma irradiation: Genome-wide expression profiling of the ROS scavenging and signal transduction pathways. *J Plant Physiol* 168:1960–1971.
40. Ricaud L, et al. (2007) ATM-mediated transcriptional and developmental responses to gamma-rays in *Arabidopsis*. *PLoS One* 2:e430.
41. Missirian V, Conklin PA, Culligan KM, Huefner ND, Britt AB (2014) High atomic weight, high-energy radiation (HZE) induces transcriptional responses shared with conventional stresses in addition to a core "DSB" response specific to clastogenic treatments. *Front Plant Sci* 5:364.
42. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z (2007) Reconstructing dynamic regulatory maps. *Mol Syst Biol* 3:74.
43. Schulz MH, et al. (2012) DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst Biol* 6:104.
44. Bourbousse C, Vegesna N, Law JA (2018) The SOG1 transcriptional activator and the MYB3R repressors control a complex gene network in response to DNA damage in Arabidopsis. Gene Expression Omnibus (GEO). Available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112773. Deposited April 5, 2018.
45. Davuluri RV, et al. (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis cis*-regulatory elements and transcription factors. *BMC Bioinformatics* 4:25.
45. Palaniswamy SK, et al. (2006) AGRIS and AtRegNet, a platform to link *cis*-regulatory elements and transcription factors into regulatory networks. *Plant Physiol* 140:818–829.
47. Yilmaz A, et al. (2011) AGRIS: The *Arabidopsis* Gene Regulatory Information Server, an update. *Nucleic Acids Res* 39:D1118–D1122.
48. O'Malley RC, et al. (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165:1280–1292.
49. Fu ZQ, Dong X (2013) Systemic acquired resistance: Turning local infection into global defense. *Annu Rev Plant Biol* 64:839–863.
50. Song J, Bent AF (2014) Microbial pathogens trigger host DNA double-strand breaks whose abundance is reduced by plant defense responses. *PLoS Pathog* 10:e1004030.
51. Shao H, Wang H, Tang X (2015) NAC transcription factors in plant multiple abiotic stress responses: Progress and prospects. *Front Plant Sci* 6:902.
52. Zhang S, et al. (2015) C-terminal domains of a histone demethylase interact with a pair of transcription factors and mediate specific chromatin association. *Cell Discov* 1:15003.
53. Chen P, et al. (2017) *Arabidopsis* R1R2R3-Myb proteins are essential for inhibiting cell division in response to DNA damage. *Nat Commun* 8:635.
54. Haga N, et al. (2011) Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in *Arabidopsis*. *Plant Physiol* 157:706–717.
55. Ito M, et al. (1998) A novel cis-acting element in promoters of plant B-type cyclin genes activates M phase-specific transcription. *Plant Cell* 10:331–341.
56. Ito M, et al. (2001) G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. *Plant Cell* 13:1891–1905.
57. Menges M, de Jager SM, Gruissem W, Murray JA (2005) Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J* 41:546–566.
58. Heitzeberg F, et al. (2004) The Rad17 homologue of *Arabidopsis* is involved in the regulation of DNA damage repair and homologous recombination. *Plant J* 38:954–968.
59. Uanschou C, et al. (2007) A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *EMBO J* 26:5061–5070.
60. Schommer C, Beven A, Lawrenson T, Shaw P, Sablowski R (2003) AHP2 is required for bivalent formation and for segregation of homologous chromosomes in *Arabidopsis* meiosis. *Plant J* 36:1–11.
61. Domenichini S, Raynaud C, Ni DA, Henry Y, Bergounioux C (2006) Atmnd1-delta1 is sensitive to gamma-irradiation and defective in meiotic DNA repair. *DNA Repair (Amst)* 5:455–464.
62. Kerzendorfer C, et al. (2006) The *Arabidopsis thaliana* MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *J Cell Sci* 119:2486–2496.
63. Dean PJ, et al. (2009) A novel ATM-dependent X-ray-inducible gene is essential for both plant meiosis and gametogenesis. *Plant J* 58:791–802.
64. Liu Y, Deng Y, Li G, Zhao J (2013) Replication factor C1 (RFC1) is required for double-strand break repair during meiotic homologous recombination in *Arabidopsis*. *Plant J* 73:154–165.
65. Wang Y, et al. (2012) The DNA replication factor RFC1 is required for interference-sensitive meiotic crossovers in *Arabidopsis thaliana*. *PLoS Genet* 8:e1003039.
66. Kobbe D, Blanck S, Focke M, Puchta H (2009) Biochemical characterization of AtRECQ3 reveals significant differences relative to other RecQ helicases. *Plant Physiol* 151:1658–1666.
67. Diaz M, Pecinka A (2018) Scaffolding for repair: Understanding molecular functions of the SMC5/6 complex. *Genes (Basel)* 9:E36.
68. Wei W, et al. (2012) A role for small RNAs in DNA double-strand break repair. *Cell* 149:101–112.
69. Shultz RW, Tatineni VM, Hanley-Bowdoin L, Thompson WF (2007) Genome-wide analysis of the core DNA replication machinery in the higher plants *Arabidopsis* and rice. *Plant Physiol* 144:1697–1714.
70. García-Ortiz MV, Ariza RR, Hoffman PD, Hays JB, Roldán-Arjona T (2004) *Arabidopsis thaliana* AtPOLK encodes a DinB-like DNA polymerase that extends mispaired primer termini and is highly expressed in a variety of tissues. *Plant J* 39:84–97.
71. Yin H, et al. (2009) Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in *Arabidopsis*. *Plant Cell* 21:386–402.
72. Wang C, Liu Z (2006) *Arabidopsis* ribonucleotide reductases are critical for cell cycle progression, DNA damage repair, and plant development. *Plant Cell* 18:350–365.
73. Clausen AR, et al. (2012) Two thymidine kinases and one multisubstrate deoxyribonucleoside kinase salvage DNA precursors in *Arabidopsis thaliana*. *FEBS J* 279:3889–3897.
74. Xu J, Zhang L, Yang DL, Li Q, He Z (2015) Thymidine kinases share a conserved function for nucleotide salvage and play an essential role in *Arabidopsis thaliana* growth and development. *New Phytol* 208:1089–1103.
75. Pedroza-García JA, Nájera-Martínez M, de la Paz Sanchez M, Plasencia J (2015) *Arabidopsis thaliana* thymidine kinase 1a is ubiquitously expressed during development and contributes to confer tolerance to genotoxic stress. *Plant Mol Biol* 87:303–315.
76. Jacobs AL, Schär P (2012) DNA glycosylases: In DNA repair and beyond. *Chromosoma* 121:1–20.
77. Leung CC, Glover JN (2011) BRCT domains: Easy as one, two, three. *Cell Cycle* 10:2461–2470.
78. Cantor SB, Nayak S (2016) FANCJ at the FORK. *Mutat Res* 788:7–11.
79. Li C, Potuschak T, Colón-Carmona A, Gutiérrez RA, Doerner P (2005) *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. *Proc Natl Acad Sci USA* 102:12978–12983.
80. De Veylder L, et al. (2001) Functional analysis of cyclin-dependent kinase inhibitors of *Arabidopsis*. *Plant Cell* 13:1653–1668.
81. Kumar N, et al. (2015) Functional conservation in the SIAMESE-RELATED family of cyclin-dependent kinase inhibitors in land plants. *Plant Cell* 27:3065–3080.
82. De Schutter K, et al. (2007) *Arabidopsis* WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. *Plant Cell* 19:211–225.
83. Vlieghe K, et al. (2005) The DP-E2F-like gene DEL1 controls the endocycle in *Arabidopsis thaliana*. *Curr Biol* 15:59–63.
84. Engeland K (2015) Simplify p53: Just an activator. *Oncotarget* 6:3–4.
85. Fischer M, Steiner L, Engeland K (2014) The transcription factor p53: Not a repressor, solely an activator. *Cell Cycle* 13:3037–3058.
86. Niida H, Shimada M, Murakami H, Nakanishi M (2010) Mechanisms of dNTP supply that play an essential role in maintaining genome integrity in eukaryotic cells. *Cancer Sci* 101:2505–2509.
87. Pillaire MJ, Bétous R, Hoffmann JS (2014) Role of DNA polymerase κ in the maintenance of genomic stability. *Mol Cell Oncol* 1:e29902.
88. Polo SE (2015) Reshaping chromatin after DNA damage: The choreography of histone proteins. *J Mol Biol* 427:626–636.
89. Fischer M (2016) p21 governs p53's repressive side. *Cell Cycle* 15:2852–2853.
90. Kobayashi K, et al. (2015) Transcriptional repression by MYB3R proteins regulates plant organ growth. *EMBO J* 34:1992–2007.
91. Roitinger E, et al. (2015) Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and rad3-related (ATR) dependent DNA damage response in *Arabidopsis thaliana*. *Mol Cell Proteomics* 14:556–571.
92. Waterworth WM, Bray CM, West CE (2015) The importance of safeguarding genome integrity in germination and seed longevity. *J Exp Bot* 66:3549–3558.
93. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
94. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.
95. Ramírez F, et al. (2016) deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44:W160–W165.
96. Kestler HA, et al. (2008) VennMaster: Area-proportional euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics* 9:67.
97. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208.
98. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. *Genome Biol* 8:R24.
99. Boyle EI, et al. (2004) GO:TermFinder—Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20:3710–3715.
100. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.
101. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
102. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
103. Mi H, et al. (2017) PANTHER version 11: Expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45:D183–D189.
104. Krishnakumar V, et al. (2015) Araport: The *Arabidopsis* information portal. *Nucleic Acids Res* 43:D1003–D1009.
105. Fischer M (2017) Census and evaluation of p53 target genes. *Oncogene* 36:3943–3956.
106. Kenzelmann Broz D, et al. (2013) Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. *Genes Dev* 27:1016–1031.

PNAS
www.pnas.org

Supplementary Information for

The SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis*

Clara Bourbousse, Neeraja Vegesna and Julie A. Law

Julie A. Law
Email: jlaw@salk.edu

This PDF file includes:

Supplementary Material and Methods
Figs. S1 to S14
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S7
Source Data 1 to 4

www.pnas.org/cgi/doi/10.1073/pnas.1810582115

Material and Methods

Plant materials

The *Arabidopsis* lines utilized include the Columbia (Col) wild-type strain, the *sog1-1* EMS point mutant provided by Dr. Anne Britt, which is in a mixed Landsberg *erecta* (L*er*)/Col background containing a *CYCB1;1 promoter::GUS* fusion (1), and the *myb3r1,3,5* triple mutant provided by Dr. Masaki Ito (2).

mRNA sequencing experiments

Plant growth conditions and gamma-irradiation: For all mRNA sequencing experiments, seeds were surface sterilized using either vapor-phase Bleach-HCl sterilization or Ethanol/SDS and sown on 1/2 MS media with 0.6% Plant agar. After 3 days of stratification at 4°C, plates were transferred into a growth chamber set at a constant temperature of 23°C and constant light (100 µmol/m^2s).

For the wild-type and *sog1* "γ-IR vs. mock DREM model" rep1 and rep2 data sets, 6-day-old seedlings were gamma-irradiated at a dose of 100 Gy using a Co60 radioactive source at a dose-rate of 10 Gy/min. Non-irradiated seedlings were grown in parallel to provide mock controls for each time point. Eight mock or irradiated seedlings were harvested at different time points from 20 min to 24 h after the beginning of the irradiation. For the wild-type and *sog1* "t vs. t_0" rep1 and rep2 data sets, 6-day-old seedlings were gamma-irradiated at a dose of 100 Gy using a Co60 radioactive source at a dose-rate of 10 Gy/min. Non-irradiated seedlings harvested at time zero (before irradiation) were used as a reference for irradiated seedlings harvested at different time points from 10 min to 24 h after the beginning of the irradiation. For the wild-type only "γ-IR vs. mock" rep1 and rep2 data sets, 6-day-old seedlings were gamma-irradiated at a dose of 100 Gy using a Co60 radioactive source at a dose-rate of 10 Gy/min. Non-irradiated seedlings were grown in parallel to provide mock controls for each time point. Six mock or irradiated seedlings were harvested at different time points from 10 min to 24 h after the beginning of the irradiation. For the wild-type and *myb3r1,3,5* "*myb3r1,3,5* vs. wt" rep1 and rep2 data sets, 6-day-old seedlings were gamma-irradiated at a dose of 100 Gy using a Co60 radioactive source at a dose-rate of 12.4 Gy/min. Six mock or irradiated seedlings were harvested 3 h after irradiation.

RNA isolation, library preparation and sequencing: For all mRNA-seq experiments, RNA was extracted from pooled seedlings using the Quick-RNA™ MiniPrep kit (Zymo Research R1055). mRNAs were then purified from 1-2 µg of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490) prior to library preparation with the NEBNext Ultra or UltraII RNA Library Prep Kit for Illumina (NEB E7530 or E7770). All sequencing was performed on an Illumina HiSeq 2500 in 50 bp single end mode.

Data processing: For all mRNA-seq experiments, Illumina reads were mapped to the TAIR10 genome using STAR (3) with the following parameters: maximum number of mismatches per read = 2, minimum intron length = 20 bp, maximum intron length = 6000 bp, minimum total length of exons = 5% of read length. A summary of the read mapping is presented in **Dataset S1**. Expression values for all genes and samples were retrieved using the analyzeRepeat.pl script from HOMER (4) with the -noadj and -condenseGenes options. These values were used to generate the PCA plots for the wild-type and *sog1* "γ-IR vs. mock DREM model" rep1 and

rep2 data sets (**SI Appendix, Fig. S1A**) in R studio using ggplot. These values were also utilized to identify differentially expressed (DE) genes based on the experimental conditions (*i.e.* γ-IR vs. mock; **Dataset S2**) or genotype (*i.e.* wild-type vs. mutant; **Dataset S5A**) using the getDiffExpression.pl script from HOMER (4) with the -DESeq2 and -repeats options and a threshold of FC≥2 with an FDR≤0.01. To identify DE genes based on both the experimental conditions and genotypes for the "wt vs. *myb3r1,3,5*" experiment (**Dataset S5B**, **C**), the analysis was performed in R studio using DESeq2 with design = batch + Condition + Group + Condition:Group. For the supplementary Datasets, gene annotations were retrieved using the Thalemine data mining tool (5).

To visualize the gene expression data, UCSC browser tracks were generated for all mRNA-seq samples using the makeUCSCfile script from HOMER (4), normalizing to ten million reads and using the following additional options: -fragLength given -style rnaseq -strand both. In addition, tracks showing the difference between the average mock and γ-IR treated samples were generated as follows using deepTools (6). First, the average expression levels of each set of biological replicates was determined from the bam files generated during the STAR mapping using the bamCompare function with the -ratio mean, -bs 20, and -normalizeUsingRPKM options. Next, the differences between the mock and γ-IR samples were determined using the bigwigCompare function with the -ratio subtract and -bs 20 options. These subtraction tracks are presented in **Fig. 1B**. Finally, heatmaps showing the expression of individual genes during the γ-IR time course were generated using the normalized $\log_2$ fold-change values generated by DESeq2 and plotted in R studio using pheatmap. Venn diagrams comparing DE gene sets were generated using VennMaster (7).

Dynamic Regulatory Events Miner (DREM) analysis

Data inputs (see **Source Data 1**): Three different types of files were generated as inputs for the DREM models. The first is a list of the 33,323 TAIR10 gene IDs, named "allgenes_ID.txt". The second is a set of gene-TF interactions. Here, two different files were generated. Initially, a file of gene-TF interactions was compiled from the AGRIS and DAP-seq databases and named "8_TF_TFname_targetGene_DAP_ampDAP_agris.txt". To generate this file, only the 4,946 gene-TF interactions from the AGRIS AtRegNet database (http://arabidopsis.med.ohio-state.edu/) (8-10) classified as "direct" and "confirmed" were included, and for the DAP-seq data (11), the DAP Peaks (FRiP ≥ 5%) were downloaded from the Plant Cistrome Database (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/browse_table_aj.php) and the peaks were assigned to all genes with a TSS within 1 kb of the peak summit in R studio using the ChIPpeakAnno package. After additional analysis of the DREM network, the SOG1 gene-TF interactions identified here and the MYB3R3 gene-TF interactions from (2) were added to generate a new file named "9_TF_TFname_targetGene_DAP_ampDAP_agris SOG1_MYB3R3_2931.txt". The third file type contains the expression values from the RNA-seq experiments. Here, two different files were generated, one for the wild-type data "wt_ratio_2395DEgenes" and another for the *sog1* data "*sog1*_ratio_2395DEgenes_rounded.txt". These files contain the $\log_2$ expression values obtained from DESeq2 output files for each of the γ-IR vs. mock time points. Only the 2,395 differentially expressed genes (FC≥2 and FDR≤0.01) complied from the wild-type and *sog1* γ-IR time courses were included. "NA" values returned by DESeq2 were replaced by blanks.

Models and parameters (see **Source Data 1**): Using the input files described above the wild-type and *sog1* DREM models were generated by a two-step process using the DREM version 2.0.3 (12, 13) as follows: For the wild-type analysis, a first model "wt_model_1" was generated using the files and parameters listed in the "wt_options_1.txt" file. The final model "wt_model_2" was then generated based on the initial wt_model_1 using the options listed in the "wt_options_2.txt" file. For the *sog1* DREM analysis, a first model "*sog1*_model_1" was generated using the files and parameters listed in the "*sog1*_options_1.txt" file. The final model "*sog1*_model_2" was then generated based on the initial *sog1*_model_1 using the options listed

in the "*sog1*_options_2.txt" file. These models were all built without using the gene-TF interaction data and utilized the Train-test framework, but they differ in the number of missing values allowed, the minimum standard deviation values and the Train-test parameters.

Motif analysis

The MEME suite (14) was used to identify motifs enriched in the promoters of genes present in the individual DREM paths. The PSP ("position-specific prior") approach (15) was used to allow MEME to perform discriminative motif discovery. Specifically, position specific priors were generated through the psp-gen function of the MEME suite using the promoter sequences from all *Arabidopsis* genes as the background set of sequences and the promoter sequences of the analyzed dataset as the primary set of sequences. In addition, a 0-order Markov model file was calculated from the background set of promoter sequences to model Markov background probabilities adjusting for single letter biases. For these analyses, promoters were defined as -500 bp to +350 bp after the TSS and when several isoforms of a gene were defined in the TAIR10 database we defined its promoter as +350 bp from the most internal TSS and -500 from the most external TSS based on all the isoforms. A maximum of 10 motifs per DREM path, ranging from 6 to 18 bp, with an occurrence of 0 to 1 per sequence were identified. All found motifs were then compared to the DAP-seq database of motifs (11) using Tomtom (16) to search for motif matches overlapping by at least 5 bp with a distance coefficient calculated by the pearson method showing an E-value <10. Only the matches with an E-value <0.01 were kept.

Gene Ontology (GO) analysis

Identification of enriched GO terms: Gene ontology enrichments for each DREM path were retrieved using the GO-TermFinder software (17) via the Princeton GO-TermFinder interface (http://go.princeton.edu/cgi-bin/GOTermFinder) in the batch mode. For this analysis, the following options were selected: "Process" for the ontology aspect, "TAIR-A. thaliana" for the annotation, and a p-value cutoff of 1, to retrieve GO enrichment values for all GO terms and allow for subsequent comparisons between DREM paths. The gene list input files and the resulting raw GO analysis data are included as part of **Source Data 2**. To reduce and visualize the GO data, the REVIGO (18) platform was utilized (http://revigo.irb.hr/) with the allowed similarity level set to "Tiny". The resulting REVIGO GO tables are included as part of **Source Data 2**.

Data processing and heatmap generation: From the REVIGO GO table all the terms not eliminated by the similarity filter (*i.e.* those that do not have null values for the Plot_X and Plot_Y columns) were compiled across either the wild-type DREM paths (W1-W11) or the *sog1* DREM paths (S1-S5) and then exact duplicates in term names were removed to generate lists of all the unique GO terms that were significantly enriched in one or more path from the wild-type or *sog1* DREM models, respectively. Using just these unique GO terms, table containing the $\log_{10}$ p-values for each term within each DREM path was generated based on the values in the original GO tables generated by Princeton GO-TermFinder. Unclustered heatmaps showing the enrichment of either all GO terms (**SI Appendix, Fig. S3** and **Fig. S6D**) or GO terms with $\log_{10}$ p-values <-1.7 in at least one DREM path (**Fig. 1C**) were generated using Treeview. Grey boxes indicate p-values >0.5, which is the cutoff employed by REVIGO.

Genevestigator Analyses

To identify additional stresses causing similar transcriptional signatures as observed in specific DREM paths, we used the Genevestigator signature tool (19). For these analyses we compared the $\log_2$ FC in the expression (γ-IR vs. mock) of genes in paths W1-W3 (3 h), W5 (1 h 30 min), W6 (20 min), and S1 (20 min), at the time points indicated in the parentheses, to the following

expression data available through Genevestigator: AT-00120 (Cold, Heat, Oxidative, Salt, Drought, Genotoxic, and wounding); AT-00107 (Response to bacterial-(LPS, HrpZ, Flg22) and the oomycete-(NPP1) derived elicitors); AT-00108 (Response to *Phytophthora infestans*); AT-00025 (Functional Genomics of Ozone Stress in Arabidopsis); AT-00197 (Response of Arabidopsis seedlings to polychromatic radiation with decreasing short-wave cut-off in the UV range); AT-00620 (Expression data from Arabidopsis leaves after gamma irradiation of 200 Gy); AT-00647 (Transcription profiling by array of Arabidopsis after treatment with hydroxyurea). The relative similarity factors, which indicate the degree of resemblance between our DREM paths and each dataset, relative to the other datasets, were retrieved and displayed as an unclustered heatmap using Treeview.

SOG1 tagging, complementation and ChIP-seq experiments

Tagging and complementation: The SOG1 promoter (*pSOG1*) and genomic sequences were amplified by PCR from Columbia genomic DNA and the 3xFLAG-BLRP sequence was amplified from a previously described plasmid (20) using the oligonucleotides detailed in **Dataset S7**. The pSOG1, SOG1 and 3xFLAG-BLRP PCR fragments were cloned using Gateway technology into pDONRP4-P1R, pDONR221 and pDONRP2R-P3, respectively and recombined into the pH7m34GW binary vector to generate a *pSOG1::SOG1-3xFLAG* construct that was transformed into *sog1* mutant plants using the floral dip method (21). Several independent, single insertion lines were selected on 1/2 MS media with hygromycin (20 µg/mL) and the SOG1-3xFLAG protein was visualized by western blotting using T_2 plants. Four hygromycin-resistant seedlings were ground in liquid nitrogen 4 hours after UV-C irradiation (3000 J/m^2). The powder was resuspended in 200 µL 2X Laemmli and boiled for 10 min. Samples were migrated on a Criterion Bis-Tris 4-12% gel (BioRad 3450123), transferred to a PVDF membrane, incubated with HRP-conjugated anti-FLAG antibodies (1:10,000; SIGMA A8592), and detected using the PierceTM ECL Plus Western Blotting Substrate (Thermo Scientific 32132).

To assess the ability of the *pSOG1::SOG1-3xFLAG* transgenic lines to complement the gene expression defects observed in *sog1* mutants, qRT-PCR experiments were conducted. For these assays, 6-day-old seedlings grown under long day conditions were transferred for 24 h in liquid MS media. Seedlings were irradiated with UV-C at a dose of 1200 J/m^2 in a Stratalinker (Stratagene). UV-C and mock treated plants were harvested 3 h after treatment. RNA isolation was performed as described above. Reverse transcription was conducted on 1 µg of RNA using random hexamers and a cDNA reverse transcription kit (High-Capacity cDNA Reverse Transcription Kit, Applied Biosystems 4368813). Expression levels of *GMI1*, *BRCA1* and *AT4G25330* genes were assessed by qPCR using a LightCycler 480 and LightCycler 480 SYBR green I Master mix (Roche 04 707 516 001). Values were normalized relative to three genes with invariant expression, *AT4G29130*, *AT2G36060*, and *AT5G13440*, and error bars represent the standard deviation of two technical replicates. Primer sequences are available in **Dataset S7**.

Chromatin immunoprecipitation: For the ChIP experiments, seeds were surface sterilized using vapor-phase Bleach-HCl sterilization and sown on 1/2 MS media with 0.6% Plant agar. After 3 days of stratification at 4°C, plates were transferred into a growth chamber set at a constant temperature of 23°C and constant light (100 µmol/m^{2}s). 10-day-old seedlings were gamma-irradiated at a dose of 75 Gy at a dose-rate of 3.75 Gy/min. 1.3 g of tissue from wild-type and SOG1-3xFLAG transgenic lines, harvested after 20 min and 1 h of γ-IR, were ground in liquid nitrogen and resuspended in 25 mL of nuclear isolation buffer (60 mM HEPES, pH 8.0, 1 M Sucrose, 5 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 0.6% Triton X-100, 0.4 mM PMSF, 1 µM pepstatin, complete mini EDTA-Free protease inhibitor (Roche 11836170001)). Proteins were then crosslinked first with 1.5 mM EGS (ethylene glycol bis(succinimidyl succinate); Thermo Fisher Scientific 21565) for 20 min at room temperature, and then with formaldehyde at a final concentration of 1% for 10 min at room temperature

mL of 2 M glycine and incubating for 10 min at room temperature. The crosslinked samples were filtered through Miracloth and centrifuged at 4,500 x g for 20 min at 4°C. Pellets were resuspended in 1 mL EB2 (0.25 M Sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 1 mM EDTA, 5 mM β-Mercaptoethanol, 0.1 mM PMSF, 1 µM pepstatin, complete mini EDTA-Free protease inhibitor (Roche 11836170001)), and centrifuged at 12,500 x g for 10 min at 4°C. Pellets were resuspended in 300 µL of nuclei lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, 0.1 mM PMSF, 1 µM pepstatin, complete mini EDTA-Free protease inhibitor (Roche 11836170001)). The chromatin was then sonicated using a Bioruptor (Diagenode) for 20 cycles (1 cycle = 30 sec on-30 sec off, on high intensity) at 4°C. The debris was pelleted by centrifugation at 9,600 x g for 10 min at room temperature. 30 µL were kept as an input. The remaining sample was diluted 10 times using ChIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl, 0.1 mM PMSF, 1 µM pepstatin, complete mini EDTA-Free protease inhibitor (Roche 11836170001)). 25 µL of M2-FLAG magnetic bead slurry (SIGMA M8823) was added and the sample was rotated for 1 h at 4°C. Beads were washed sequentially in 1 mL of low salt buffer (150 mM NaCl, 0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0), 1 mL of high salt buffer (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0), 1 mL LiCl buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) and 1 mL of TE (10 mM Tris-HCl, pH8.0, and 1 mM EDTA). The chromatin elution was performed in 300 µL SDS elution buffer (1% SDS, 0.1 M $NaHCO_3$) at 65°C and the crosslinking was reversed by adding 12 µL of 5 M NaCl and 30 µL of 1 M DTT and incubating the samples overnight at 65°C. Digestion with Proteinase K was performed at 45°C for 1 h. DNA was purified by phenol-chloroform-IAA extraction, and precipitated with ethanol. Pellets of both inputs and IPs were resuspended in 30 µL of TE, pH 8.0.

Library preparation and sequencing: Libraries were prepared from 25 µL of the IP samples or 1 µg of the input samples using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645). Sequencing was performed on an Illumina HiSeq 2500 in 50 bp single-end mode.

ChIP mapping and data analysis: Reads were mapped to the TAIR10 genome using bowtie2 (22) with the default parameters and mapping statistics are listed in **Dataset S4A**. UCSC browser tracks were generated using the HOMER suite of NGS tools (4) and normalized to ten million reads. SOG1 peaks at 20 min or 1 h were called independently relative to two controls (the wt_IP and input samples) using the HOMER findPeaks tool (4) and only peaks identified relative to both controls were kept. For peak calling, a Fold Change of 2.25 (-F2.25) and an FDR>0.001 were required and the following additional findPeaks parameters were used (-style factor -region -L2 -ntagThreshold 42 -minDist187 -center). As the peaks identified at the 20 min and 1 h time points showed significant overlaps, these lists were merged, resulting in a set of 310 peaks, three of which were removed due to high background (**Dataset S4B**). The remaining 307 peaks were then assigned to gene targets as follows: First, the four nearest neighboring genes to each peak were identified using bedtools closest (23) with the -k4 and -D b options. Second, peaks were assigned to the nearest TSS with preference going to genes with altered expression upon exposure to gamma-irradiation when two genes shared a common promoter. With the exception of 7 peaks that were assigned to two genes, each peak was assigned to a single gene. While most genes were associated with just one SOG1 peak, 4 genes had two distinct SOG1 peaks. Thus, the 307 peaks (**Dataset S4B**) were assigned to 310 genes (**Dataset S4C**, **D**). Gene annotations were retrieved using the Thalemine data mining tool (5) and overlaps between SOG1 target genes identified here or in (24) were generated using VennMaster (7).

Peak locations relative to genome features (promoters, introns, exons etc.) were determined (4), where the promoter regions are defined as -1 kb and +100 bp relative to the TSS. Motifs under the SOG1 peaks were determined using

the MEME tool (25) from the MEME suite (14) with the following options (-nostatus -maxsize 7500000 -nmotifs 10 -minw 6 -maxw 18 -revcomp -psp -bfile).

Heatmaps and Metaplots were generated using the deepTools suite (6). The ChIP data was processed as follows. First the data was either normalized using bamCoverage with the following options (-b *.bam -v -bs 10 -normalizeUsingRPKM -extendReads 300) to generate normalized bigwig (.bw) files for the input and IP samples, or compared (SOG1 ChIP vs. Input or SOG1 ChIP vs. wild-type ChIP) using bamCompare with the following options (-b1 sample1.bam -b2 sample2.bam -v -bs 10 -ratio=log2 -extendReads 300 -scaleFactorsMethod SES). For **Fig. 3A**, a data matrix was generated from the bamCompare files using computeMatrix with the following options (-referencePoint=center -S Comparison1.bw Comparison2.bw etc -R ChIPpeaks.bed -a 3000 -b 3000 -bs=10 -sortRegions=no) and the heatmap was generated using plotHeatmap with the following options (-m Matrix.gz -refPointLabel=Center -sortUsingSamples 3 4 -zMin -2 -zMax 2.5). For **SI Appendix, Fig. S10D**, the profile plot was generated using plotProfile with the following options (-m Matrix.gz -refPointLabel=Center -perGroup -plotType=se) with the same matrix used for **Fig. 3A**. For **SI Appendix, Fig. S10G**, a new data matrix was generated using computeMatrix with the following options (scale-regions -S SOG1_ChIP_1h_vsWt_log2ratio.bw -R W1_genes_TSS_to_TTS.bed . . . W11_genes_TSS_to_TTS.bed -a 3000 -b 3000 -bs=10 -sortRegions keep) and the heatmap was generated using plotHeatmap with the following options (-m Matrix.gz -zMin -1.25 -zMax 2.5 -sortRegions no). Heatmaps showing the expression of SOG1 target genes during the γ-IR time course (**Fig. 3B**) were generated using the $\log_2$ fold-change values generated by DESeq2 and plotted in R studio using pheatmap.

MYB3R3 ChIP-seq analysis

Reads and peak coordinates from (2) were mined from GEO (GSE60554). The closest TAIR10 gene to each peak was determined using annotatepeaks.pl from HOMER (4). Mapping of the ChIP-seq reads and analysis of ChIP enrichment profiles using deepTools were as described for the SOG1 ChIP-seq. Venn diagrams were generated using VennMaster (7) based on known G2/M expressed genes (26, 27) or previously defined MYB3R3 ChIP peaks (2).

Cytoscape network construction

For the functional categorization of the SOG1 targets (**Fig. 4**), we assigned 141 genes to relevant functional categories by combining information from the biological processes GO analysis and the TAIR10 annotations for each gene, including molecular functions and the presence of protein domains. A network for these 141 genes was processed with the Cytoscape 3.4.0 software (**Source Data 4**), including the $\log_2$ Fold Change in expression 3 h after γ-IR. The SOG1 targets underlined in blue represent those that have a human or mouse ortholog, identified using the PANTHER (28) and/or Thalemine tools (5), that were shown to be targeted and up-regulated by p53 based on 13 genome-wide studies in humans (29) or a single study in primary mouse embryo fibroblasts study (30). For the TF network downstream of SOG1 (**SI Appendix, Fig. S11** and **Source Data 4**), interaction data for the 33 TFs downstream of SOG1 was fetched from the AGRIS (8-10) and -seq (11) databases. Only the interactions that target TFs present in the DREM model were kept. The resulting interaction file was used to generate a network with the Cytoscape 3.4.0 software, including the $\log_2$ Fold Change in expression 3 h after γ-IR.

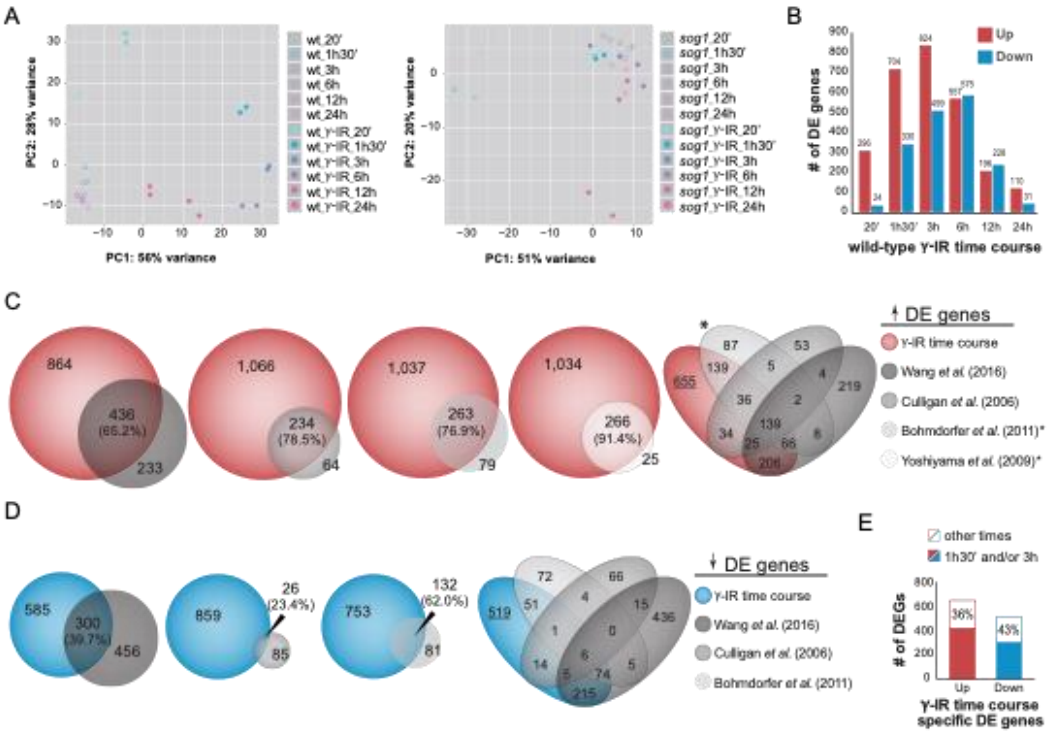

A
PC2: 28% variance
PC1: 56% variance
wt_20'
wt_1h30'
wt_3h
wt_6h
wt_12h
wt_24h
wt_γ-IR_20'
wt_γ-IR_1h30'
wt_γ-IR_3h
wt_γ-IR_6h
wt_γ-IR_12h
wt_γ-IR_24h
PC2: 20% variance
PC1: 51% variance
sogf_20'
sogf_1h30'
sogf_3h
sogf_6h
sogf_12h
sogf_24h
sogf_γ-IR_20'
sogf_γ-IR_1h30'
sogf_γ-IR_3h
sogf_γ-IR_6h
sogf_γ-IR_12h
sogf_γ-IR_24h
B
# of DE genes
Up
Down
20'
1h30'
3h
6h
12h
24h
wild-type γ-IR time course
C
DE genes
γ-IR time course
Wang et al. (2016)
Culligan et al. (2006)
Bohmdorfer et al. (2011)*
Yoshiyama et al. (2009)*
D
DE genes
γ-IR time course
Wang et al. (2016)
Culligan et al. (2006)
Bohmdorfer et al. (2011)
E
other times
1h30' and/or 3h
# of DEGs
Up
Down
γ-IR time course specific DE genes

Figure S1

Fig. S1. **Temporal profiling of gene expression changes after γ-IR reveals additional DNA damage responsive genes.** (**A**) Principle component analysis (PCA) plots showing the groupings of data points from the wild-type (wt) or *sog1* γ-IR time courses. The two biological replicates for each sample are represented using matched colors. (**B**) Bar chart showing the number of differentially expressed (DE) genes identified at each time point in the wt γ-IR time course, see **Dataset S2A**. (**C** and **D**) Scaled Venn diagrams showing the overlaps between up- or down-regulated genes, respectively, in response to γ-IR as identified in this study (γ-IR time course 20 min-24 h; 100 Gy; 6-day-old seedlings; RNA-seq with FC≥2 and FDR≤0.01), in (31) (3 h; 80 Gy; 14-day-old seedlings; RNA-seq with FC≥2 and q-value ≤0.05), in (32) (1 h 30 min; 100 Gy; 5-day-old seedlings; Affymetrix array with FC≥2, PPDE >0.95, and q-value <0.05), in (33) (1 h 30 min; 100 Gy; 5-day-old seedlings; Affymetrix array with q-value <0.05), or in (1) (1 h 30 min; 100 Gy; 5-day-old seedlings; Affymetrix array with only up-regulated genes provided). In cases where the array probes matched more than one gene, both were included. The percentage of genes from each previous data set captured in the γ-IR time course are indicated in parentheses and the number of genes uniquely identified in the γ-IR time course are underlined. For the up-regulated genes, the asterisk marks the union of two data sets that were grouped together for the final Venn diagram. Data from (34) showed much lower overlaps, perhaps due to the use of CATMA arrays, and thus were not included. (**E**) Bar graph indicating the number of γ-IR time course specific DE genes that are differentially expressed either at the previously assessed time points (filled) or only at other time points (unfilled, with the percentage of the total indicated).

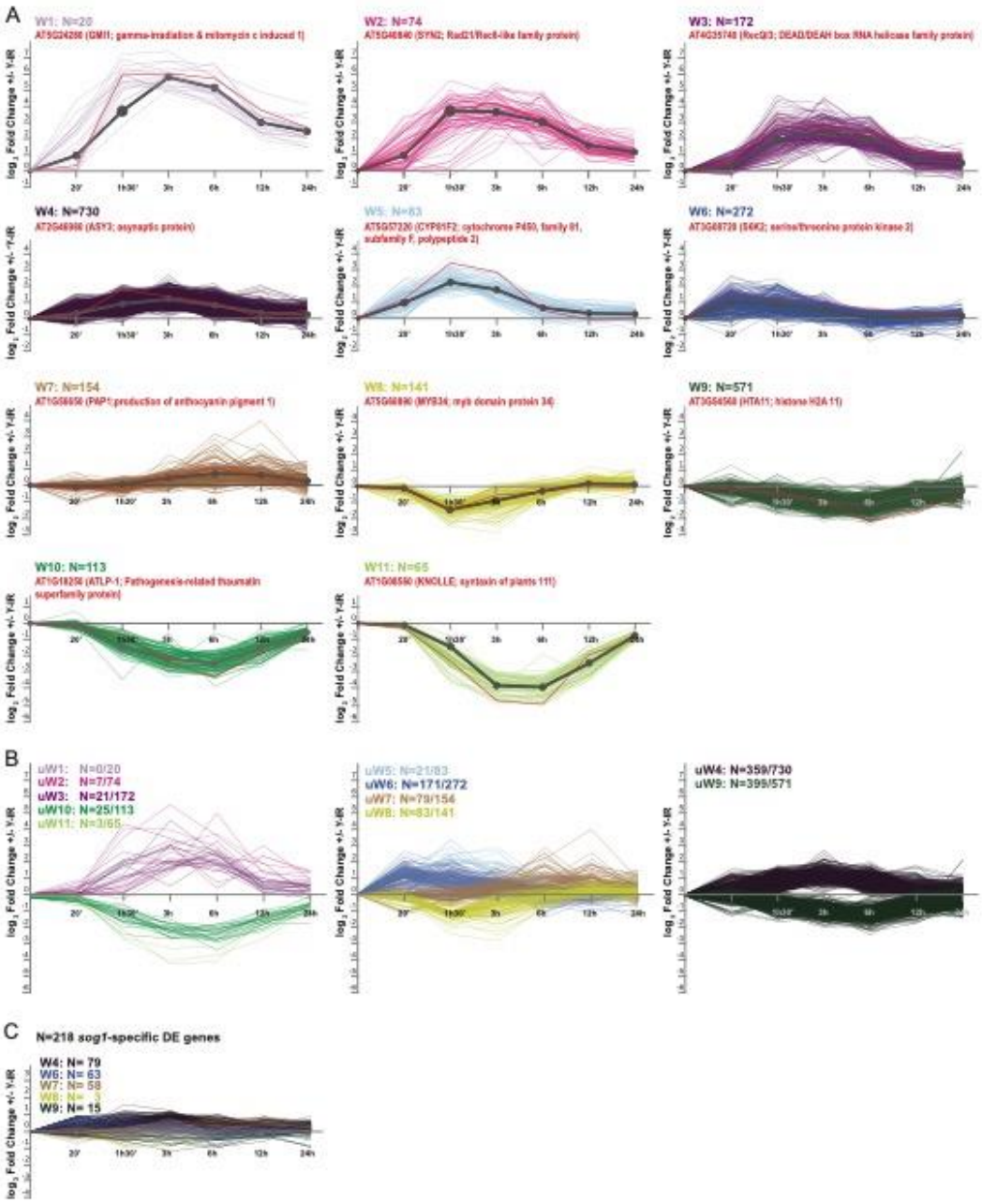

A
W1: N=20
AT5G24280 (GMI1; gamma-irradiation & mitomycin c induced 1)
W2: N=74
AT5G46840 (SYN2; Rad21/Rec8-like family protein)
W3: N=172
AT4G35740 (RecQl3; DEAD/DEAH box RNA helicase family protein)
W4: N=730
AT2G46990 (ASY3; asynaptic protein)
W5: N=83
AT5G57220 (CYP81F2; cytochrome P450, family 81, subfamily F, polypeptide 2)
W6: N=272
AT3G08720 (SIK2; serine/threonine protein kinase 2)
W7: N=154
AT1G66650 (PAP1;production of anthocyanin pigment 1)
W8: N=141
AT5G66890 (MYB34; myb domain protein 34)
W9: N=571
AT3G54560 (HTA11; histone H2A 11)
W10: N=113
AT1G18250 (ATLP-1; Pathogenesis-related thaumatin superfamily protein)
W11: N=65
AT1G08560 (KNOLLE; syntaxin of plants 111)
log2 Fold Change +/- Y-IR
B
uW1: N=0/20
uW2: N=7/74
uW3: N=21/172
uW10: N=25/113
uW11: N=3/65
uW5: N=21/83
uW6: N=171/272
uW7: N=79/154
uW8: N=83/141
uW4: N=359/730
uW9: N=399/571
C
N=218 sog1-specific DE genes
W4: N= 79
W6: N= 63
W7: N= 58
W8: N= 1
W9: N= 15

Figure S2

Fig. S2. **Gene expression patterns for wild-type DREM paths.** (**A**) Expression profiles of the individual genes in each path (W1-W11). The paths shown in **Fig. 1A** are shown in grey. The representative genes shown in **Fig. 1B** are shown in red and described above each set of genes. (**B**) Expression profiles of DE genes uniquely identified in the γ-IR time course. The DREM path and fraction of unique genes are indicated (uW#). (**C**) Expression profiles of the 218 *sog1*-specific DE genes included in the wild-type DREM model, colored based on the path in which they reside. In all panels, the *y* axis indicates the $\log_2$ Fold Change in expression in response to γ-IR and the *x* axis indicates the time in minutes (') and/or hours (h).

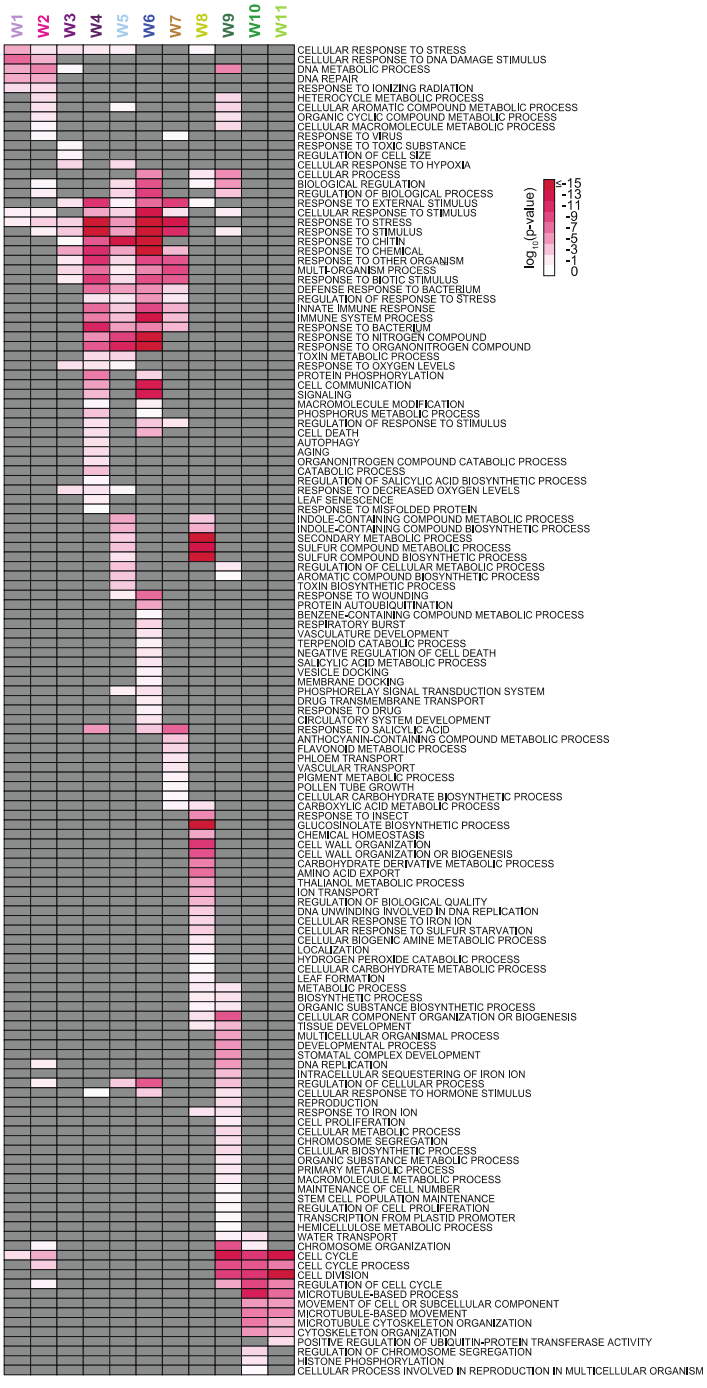

W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
W11
CELLULAR RESPONSE TO STRESS
CELLULAR RESPONSE TO DNA DAMAGE STIMULUS
DNA METABOLIC PROCESS
DNA REPAIR
RESPONSE TO IONIZING RADIATION
HETEROCYCLE METABOLIC PROCESS
CELLULAR AROMATIC COMPOUND METABOLIC PROCESS
ORGANIC CYCLIC COMPOUND METABOLIC PROCESS
CELLULAR MACROMOLECULE METABOLIC PROCESS
RESPONSE TO VIRUS
RESPONSE TO TOXIC SUBSTANCE
REGULATION OF CELL SIZE
CELLULAR RESPONSE TO HYPOXIA
CELLULAR PROCESS
BIOLOGICAL REGULATION
REGULATION OF BIOLOGICAL PROCESS
RESPONSE TO EXTERNAL STIMULUS
CELLULAR RESPONSE TO STIMULUS
RESPONSE TO STRESS
RESPONSE TO STIMULUS
RESPONSE TO CHITIN
RESPONSE TO CHEMICAL
RESPONSE TO OTHER ORGANISM
MULTI-ORGANISM PROCESS
RESPONSE TO BIOTIC STIMULUS
DEFENSE RESPONSE TO BACTERIUM
REGULATION OF RESPONSE TO STRESS
INNATE IMMUNE RESPONSE
IMMUNE SYSTEM PROCESS
RESPONSE TO BACTERIUM
RESPONSE TO NITROGEN COMPOUND
RESPONSE TO ORGANONITROGEN COMPOUND
TOXIN METABOLIC PROCESS
RESPONSE TO OXYGEN LEVELS
PROTEIN PHOSPHORYLATION
CELL COMMUNICATION
SIGNALING
MACROMOLECULE MODIFICATION
PHOSPHORUS METABOLIC PROCESS
REGULATION OF RESPONSE TO STIMULUS
CELL DEATH
AUTOPHAGY
AGING
ORGANONITROGEN COMPOUND CATABOLIC PROCESS
CATABOLIC PROCESS
REGULATION OF SALICYLIC ACID BIOSYNTHETIC PROCESS
RESPONSE TO DECREASED OXYGEN LEVELS
LEAF SENESCENCE
RESPONSE TO MISFOLDED PROTEIN
INDOLE-CONTAINING COMPOUND METABOLIC PROCESS
INDOLE-CONTAINING COMPOUND BIOSYNTHETIC PROCESS
SECONDARY METABOLIC PROCESS
SULFUR COMPOUND METABOLIC PROCESS
SULFUR COMPOUND BIOSYNTHETIC PROCESS
REGULATION OF CELLULAR METABOLIC PROCESS
AROMATIC COMPOUND BIOSYNTHETIC PROCESS
TOXIN BIOSYNTHETIC PROCESS
RESPONSE TO WOUNDING
PROTEIN AUTOUBIQUITINATION
BENZENE-CONTAINING COMPOUND METABOLIC PROCESS
RESPIRATORY BURST
VASCULATURE DEVELOPMENT
TERPENOID CATABOLIC PROCESS
NEGATIVE REGULATION OF CELL DEATH
SALICYLIC ACID METABOLIC PROCESS
VESICLE DOCKING
MEMBRANE DOCKING
PHOSPHORELAY SIGNAL TRANSDUCTION SYSTEM
DRUG TRANSMEMBRANE TRANSPORT
RESPONSE TO DRUG
CIRCULATORY SYSTEM DEVELOPMENT
RESPONSE TO SALICYLIC ACID
ANTHOCYANIN-CONTAINING COMPOUND METABOLIC PROCESS
FLAVONOID METABOLIC PROCESS
PHLOEM TRANSPORT
VASCULAR TRANSPORT
PIGMENT METABOLIC PROCESS
POLLEN TUBE GROWTH
CELLULAR CARBOHYDRATE BIOSYNTHETIC PROCESS
CARBOXYLIC ACID METABOLIC PROCESS
RESPONSE TO INSECT
GLUCOSINOLATE BIOSYNTHETIC PROCESS
CHEMICAL HOMEOSTASIS
CELL WALL ORGANIZATION
CELL WALL ORGANIZATION OR BIOGENESIS
CARBOHYDRATE DERIVATIVE METABOLIC PROCESS
AMINO ACID EXPORT
THALIANOL METABOLIC PROCESS
ION TRANSPORT
REGULATION OF BIOLOGICAL QUALITY
DNA UNWINDING INVOLVED IN DNA REPLICATION
CELLULAR RESPONSE TO IRON ION
CELLULAR RESPONSE TO SULFUR STARVATION
CELLULAR BIOGENIC AMINE METABOLIC PROCESS
LOCALIZATION
HYDROGEN PEROXIDE CATABOLIC PROCESS
CELLULAR CARBOHYDRATE METABOLIC PROCESS
LEAF FORMATION
METABOLIC PROCESS
BIOSYNTHETIC PROCESS
ORGANIC SUBSTANCE BIOSYNTHETIC PROCESS
CELLULAR COMPONENT ORGANIZATION OR BIOGENESIS
TISSUE DEVELOPMENT
MULTICELLULAR ORGANISMAL PROCESS
DEVELOPMENTAL PROCESS
STOMATAL COMPLEX DEVELOPMENT
DNA REPLICATION
INTRACELLULAR SEQUESTERING OF IRON ION
REGULATION OF CELLULAR PROCESS
CELLULAR RESPONSE TO HORMONE STIMULUS
REPRODUCTION
RESPONSE TO IRON ION
CELL PROLIFERATION
CELLULAR METABOLIC PROCESS
CHROMOSOME SEGREGATION
CELLULAR BIOSYNTHETIC PROCESS
ORGANIC SUBSTANCE METABOLIC PROCESS
PRIMARY METABOLIC PROCESS
MACROMOLECULE METABOLIC PROCESS
MAINTENANCE OF CELL NUMBER
STEM CELL POPULATION MAINTENANCE
REGULATION OF CELL PROLIFERATION
TRANSCRIPTION FROM PLASTID PROMOTER
HEMICELLULOSE METABOLIC PROCESS
WATER TRANSPORT
CHROMOSOME ORGANIZATION
CELL CYCLE
CELL CYCLE PROCESS
CELL DIVISION
REGULATION OF CELL CYCLE
MICROTUBULE-BASED PROCESS
MOVEMENT OF CELL OR SUBCELLULAR COMPONENT
MICROTUBULE-BASED MOVEMENT
MICROTUBULE CYTOSKELETON ORGANIZATION
CYTOSKELETON ORGANIZATION
POSITIVE REGULATION OF UBIQUITIN-PROTEIN TRANSFERASE ACTIVITY
REGULATION OF CHROMOSOME SEGREGATION
HISTONE PHOSPHORYLATION
CELLULAR PROCESS INVOLVED IN REPRODUCTION IN MULTICELLULAR ORGANISM
log10(p-value)
<=15
-13
-11
-9
-7
-5
-3
-1
0

Figure S3

Fig. S3. **Enriched GO terms for the wild-type DREM model.** GO enrichments for the genes in each wild-type DREM path after REVIGO similarity filtering (18). Grey indicates a p-value >0.5. See also **Source Data 2**.

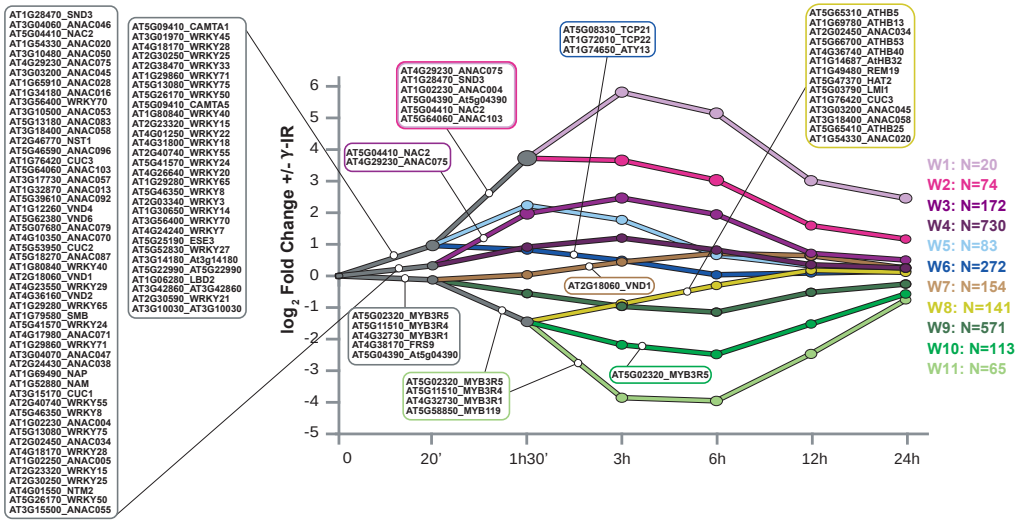

AT1G28470_SND3
AT3G04060_ANAC046
AT5G04410_NAC2
AT1G54330_ANAC020
AT3G10480_ANAC050
AT4G29230_ANAC075
AT3G03200_ANAC045
AT1G65910_ANAC028
AT1G34180_ANAC016
AT3G56400_WRKY70
AT3G10500_ANAC053
AT5G13180_ANAC083
AT3G18400_ANAC058
AT2G46770_NST1
AT5G46590_ANAC096
AT1G76420_CUC3
AT5G64060_ANAC103
AT3G17730_ANAC057
AT1G32870_ANAC013
AT5G39610_ANAC092
AT1G12260_VND4
AT5G62380_VND6
AT5G07680_ANAC079
AT4G10350_ANAC070
AT5G53950_CUC2
AT5G18270_ANAC087
AT1G80840_WRKY40
AT2G18060_VND1
AT4G23550_WRKY29
AT4G36160_VND2
AT1G29280_WRKY65
AT1G79580_SMB
AT5G41570_WRKY24
AT4G17980_ANAC071
AT1G29860_WRKY71
AT3G04070_ANAC047
AT2G24430_ANAC038
AT1G69490_NAP
AT1G52880_NAM
AT3G15170_CUC1
AT2G40740_WRKY55
AT5G46350_WRKY8
AT1G02230_ANAC004
AT5G13080_WRKY75
AT2G02450_ANAC034
AT4G18170_WRKY28
AT1G02250_ANAC005
AT2G23320_WRKY15
AT2G30250_WRKY25
AT4G01550_NTM2
AT5G26170_WRKY50
AT3G15500_ANAC055
AT5G09410_CAMTA1
AT3G01970_WRKY45
AT4G18170_WRKY28
AT2G30250_WRKY25
AT2G38470_WRKY33
AT1G29860_WRKY71
AT5G13080_WRKY75
AT5G26170_WRKY50
AT5G09410_CAMTA5
AT1G80840_WRKY40
AT2G23320_WRKY15
AT4G01250_WRKY22
AT4G31800_WRKY18
AT2G40740_WRKY55
AT5G41570_WRKY24
AT4G26640_WRKY20
AT1G29280_WRKY65
AT5G46350_WRKY8
AT2G03340_WRKY3
AT1G30650_WRKY14
AT3G56400_WRKY70
AT4G24240_WRKY7
AT5G25190_ESE3
AT5G52830_WRKY27
AT3G14180_At3g14180
AT5G22990_AT5G22990
AT1G06280_LBD2
AT3G42860_AT3G42860
AT2G30590_WRKY21
AT3G10030_AT3G10030
log2 Fold Change +/- Y-IR
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
0
20'
1h30'
3h
6h
12h
24h
AT5G08330_TCP21
AT1G72010_TCP22
AT1G74650_ATY13
AT4G29230_ANAC075
AT1G28470_SND3
AT1G02230_ANAC004
AT5G04390_At5g04390
AT5G04410_NAC2
AT5G64060_ANAC103
AT5G65310_ATHB5
AT1G69780_ATHB13
AT2G02450_ANAC034
AT5G66700_ATHB53
AT4G36740_ATHB40
AT1G14687_AtHB32
AT1G49480_REM19
AT5G47370_HAT2
AT5G03790_LMI1
AT1G76420_CUC3
AT3G03200_ANAC045
AT3G18400_ANAC058
AT5G65410_ATHB25
AT1G54330_ANAC020
AT5G04410_NAC2
AT4G29230_ANAC075
AT2G18060_VND1
AT5G02320_MYB3R5
AT5G11510_MYB3R4
AT4G32730_MYB3R1
AT4G38170_FRS9
AT5G04390_At5g04390
AT5G02320_MYB3R5
AT5G11510_MYB3R4
AT4G32730_MYB3R1
AT5G58850_MYB119
AT5G02320_MYB3R5
W1: N=20
W2: N=74
W3: N=172
W4: N=730
W5: N=83
W6: N=272
W7: N=154
W8: N=141
W9: N=571
W10: N=113
W11: N=65

Figure S4

Fig. S4. **Transcriptional regulation of the wild-type DREM paths.** The wild-type DREM model, as described in **Fig. 1A**, showing all TFs assigned to each path (ranked high to low) using the "Path Significance Conditional on Split" option with a score threshold of 0.001 and a minimum split of 20%. The *y* axis indicates the $\log_2$ Fold Change in expression in response to γ-IR and the *x* axis indicates the time in minutes (') and/or hours (h).

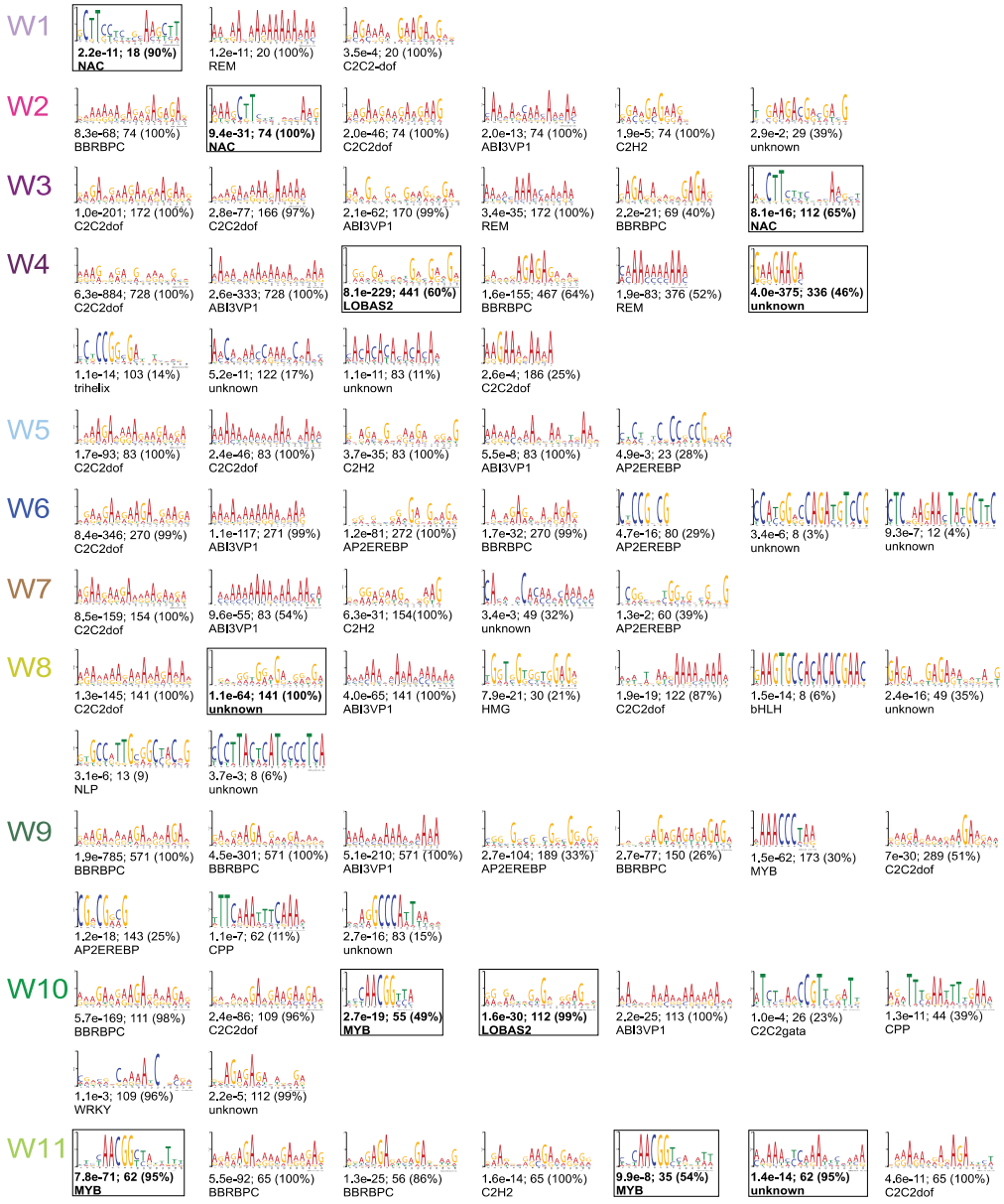

W1
2.2e-11; 18 (90%)
NAC
1.2e-11; 20 (100%)
REM
3.5e-4; 20 (100%)
C2C2-dof
W2
8.3e-68; 74 (100%)
BBRBPC
9.4e-31; 74 (100%)
NAC
2.0e-46; 74 (100%)
C2C2dof
2.0e-13; 74 (100%)
ABI3VP1
1.9e-5; 74 (100%)
C2H2
2.9e-2; 29 (39%)
unknown
W3
1.0e-201; 172 (100%)
C2C2dof
2.8e-77; 166 (97%)
C2C2dof
2.1e-62; 170 (99%)
ABI3VP1
3.4e-35; 172 (100%)
REM
2.2e-21; 69 (40%)
BBRBPC
8.1e-16; 112 (65%)
NAC
W4
6.3e-884; 728 (100%)
C2C2dof
2.6e-333; 728 (100%)
ABI3VP1
8.1e-229; 441 (60%)
LOBAS2
1.6e-155; 467 (64%)
BBRBPC
1.9e-83; 376 (52%)
REM
4.0e-375; 336 (46%)
unknown
1.1e-14; 103 (14%)
trihelix
5.2e-11; 122 (17%)
unknown
1.1e-11; 83 (11%)
unknown
2.6e-4; 186 (25%)
C2C2dof
W5
1.7e-93; 83 (100%)
C2C2dof
2.4e-46; 83 (100%)
C2C2dof
3.7e-35; 83 (100%)
C2H2
5.5e-8; 83 (100%)
ABI3VP1
4.9e-3; 23 (28%)
AP2EREBP
W6
8.4e-346; 270 (99%)
C2C2dof
1.1e-117; 271 (99%)
ABI3VP1
1.2e-81; 272 (100%)
AP2EREBP
1.7e-32; 270 (99%)
BBRBPC
4.7e-16; 80 (29%)
AP2EREBP
3.4e-6; 8 (3%)
unknown
9.3e-7; 12 (4%)
unknown
W7
8.5e-159; 154 (100%)
C2C2dof
9.6e-55; 83 (54%)
ABI3VP1
6.3e-31; 154(100%)
C2H2
3.4e-3; 49 (32%)
unknown
1.3e-2; 60 (39%)
AP2EREBP
W8
1.3e-145; 141 (100%)
C2C2dof
1.1e-64; 141 (100%)
unknown
4.0e-65; 141 (100%)
ABI3VP1
7.9e-21; 30 (21%)
HMG
1.9e-19; 122 (87%)
C2C2dof
1.5e-14; 8 (6%)
bHLH
2.4e-16; 49 (35%)
unknown
3.1e-6; 13 (9)
NLP
3.7e-3; 8 (6%)
unknown
W9
1.9e-785; 571 (100%)
BBRBPC
4.5e-301; 571 (100%)
BBRBPC
5.1e-210; 571 (100%)
ABI3VP1
2.7e-104; 189 (33%)
AP2EREBP
2.7e-77; 150 (26%)
BBRBPC
1.5e-62; 173 (30%)
MYB
7e-30; 289 (51%)
C2C2dof
1.2e-18; 143 (25%)
AP2EREBP
1.1e-7; 62 (11%)
CPP
2.7e-16; 83 (15%)
unknown
W10
5.7e-169; 111 (98%)
BBRBPC
2.4e-86; 109 (96%)
C2C2dof
2.7e-19; 55 (49%)
MYB
1.6e-30; 112 (99%)
LOBAS2
2.2e-25; 113 (100%)
ABI3VP1
1.0e-4; 26 (23%)
C2C2gata
1.3e-11; 44 (39%)
CPP
1.1e-3; 109 (96%)
WRKY
2.2e-5; 112 (99%)
unknown
W11
7.8e-71; 62 (95%)
MYB
5.5e-92; 65 (100%)
BBRBPC
1.3e-25; 56 (86%)
BBRBPC
1.6e-14; 65 (100%)
C2H2
9.9e-8; 35 (54%)
MYB
1.4e-14; 62 (95%)
unknown
4.6e-11; 65 (100%)
C2C2dof

Figure S5

Fig. S5. **Enriched motifs identified from the wild-type DREM model.** Up to 10 motifs in the promoter regions (-500 to +350 bp around the transcription start site (TSS)) of the genes in each DREM path were identified using the MEME tool (25) from the MEME suite (14). Below each motif, an E-value representing the significance of the motif is indicated along with the number and percentage of genes per path with the motif. In addition, the most similar TF family match for each motif is indicated, where unknown = no match with an E-value <1e-2. Motifs that fulfill the following criteria are boxed and labeled in bold: First, they must be largely specific to a given path and not be low complexity, which excludes the BBRBPC, ABI3VP1, C2C2dof, REM, C2H2 and AP2EREBP motifs. Second, they must also be present in >45% of the genes in a given path and have an E-value <1e-5. See also **Source Data 3**.

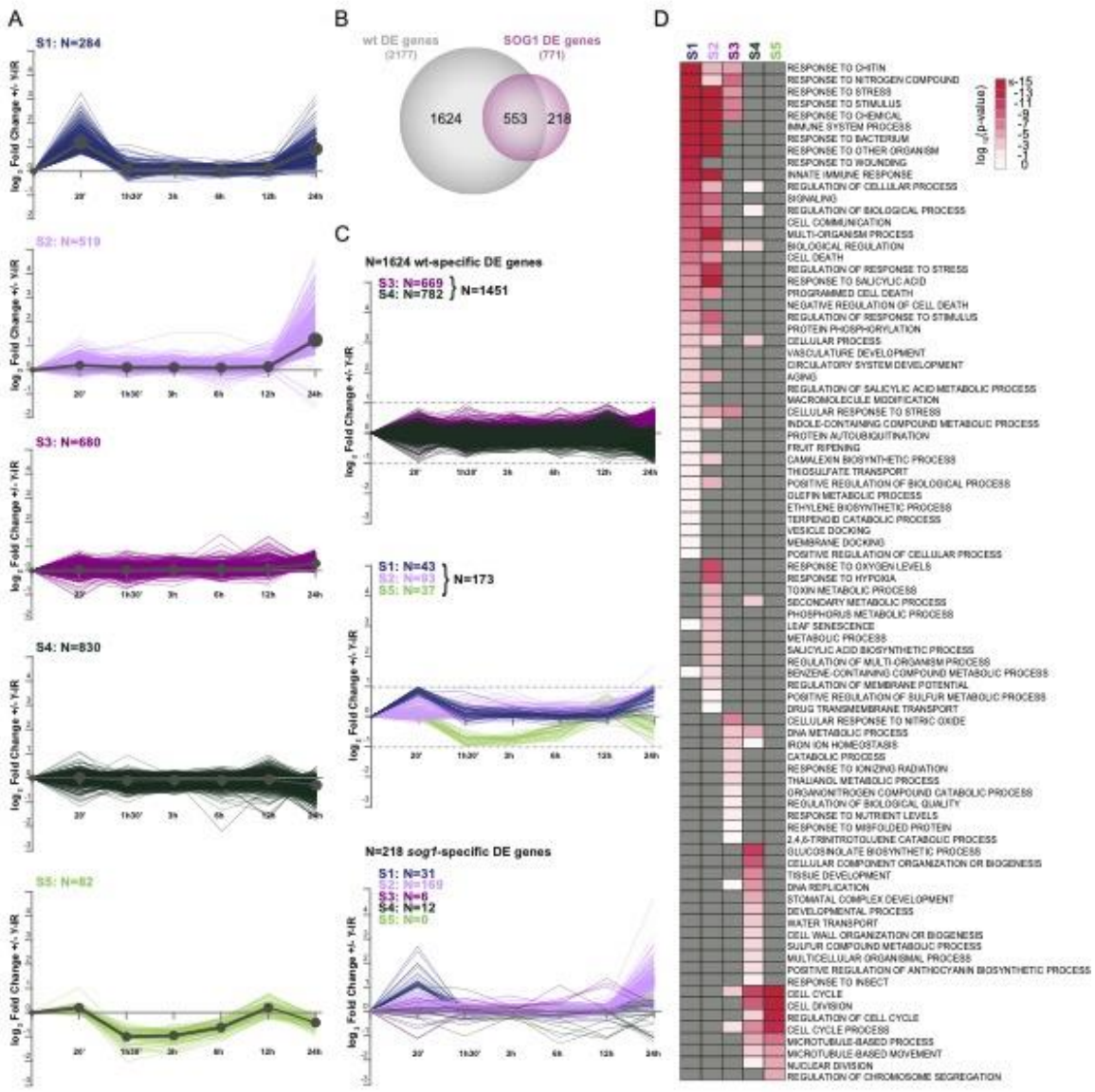

A
S1: N=284
S2: N=519
S3: N=680
S4: N=830
S5: N=82
B
wt DE genes (2177)
SOG1 DE genes (771)
1624
553
218
C
N=1624 wt-specific DE genes
S3: N=669
S4: N=782
N=1451
S1: N=43
S2: N=93
S5: N=37
N=173
N=218 sog1-specific DE genes
S1: N=31
S2: N=169
S3: N=6
S4: N=12
S5: N=0
D
S1
S2
S3
S4
S5
RESPONSE TO CHITIN
RESPONSE TO NITROGEN COMPOUND
RESPONSE TO STRESS
RESPONSE TO STIMULUS
RESPONSE TO CHEMICAL
IMMUNE SYSTEM PROCESS
RESPONSE TO BACTERIUM
RESPONSE TO OTHER ORGANISM
RESPONSE TO WOUNDING
INNATE IMMUNE RESPONSE
REGULATION OF CELLULAR PROCESS
SIGNALING
REGULATION OF BIOLOGICAL PROCESS
CELL COMMUNICATION
MULTI-ORGANISM PROCESS
BIOLOGICAL REGULATION
CELL DEATH
REGULATION OF RESPONSE TO STRESS
RESPONSE TO SALICYLIC ACID
PROGRAMMED CELL DEATH
NEGATIVE REGULATION OF CELL DEATH
REGULATION OF RESPONSE TO STIMULUS
PROTEIN PHOSPHORYLATION
CELLULAR PROCESS
VASCULATURE DEVELOPMENT
CIRCULATORY SYSTEM DEVELOPMENT
AGING
REGULATION OF SALICYLIC ACID METABOLIC PROCESS
MACROMOLECULE MODIFICATION
CELLULAR RESPONSE TO STRESS
INDOLE-CONTAINING COMPOUND METABOLIC PROCESS
PROTEIN AUTOUBIQUITINATION
FRUIT RIPENING
CAMALEXIN BIOSYNTHETIC PROCESS
THIOSULFATE TRANSPORT
POSITIVE REGULATION OF BIOLOGICAL PROCESS
OLEFIN METABOLIC PROCESS
ETHYLENE BIOSYNTHETIC PROCESS
TERPENOID CATABOLIC PROCESS
VESICLE DOCKING
MEMBRANE DOCKING
POSITIVE REGULATION OF CELLULAR PROCESS
RESPONSE TO OXYGEN LEVELS
RESPONSE TO HYPOXIA
TOXIN METABOLIC PROCESS
SECONDARY METABOLIC PROCESS
PHOSPHORUS METABOLIC PROCESS
LEAF SENESCENCE
METABOLIC PROCESS
SALICYLIC ACID BIOSYNTHETIC PROCESS
REGULATION OF MULTI-ORGANISM PROCESS
BENZENE-CONTAINING COMPOUND METABOLIC PROCESS
REGULATION OF MEMBRANE POTENTIAL
POSITIVE REGULATION OF SULFUR METABOLIC PROCESS
DRUG TRANSMEMBRANE TRANSPORT
CELLULAR RESPONSE TO NITRIC OXIDE
DNA METABOLIC PROCESS
IRON ION HOMEOSTASIS
CATABOLIC PROCESS
RESPONSE TO IONIZING RADIATION
THALIANOL METABOLIC PROCESS
ORGANONITROGEN COMPOUND CATABOLIC PROCESS
REGULATION OF BIOLOGICAL QUALITY
RESPONSE TO NUTRIENT LEVELS
RESPONSE TO MISFOLDED PROTEIN
2,4,6-TRINITROTOLUENE CATABOLIC PROCESS
GLUCOSINOLATE BIOSYNTHETIC PROCESS
CELLULAR COMPONENT ORGANIZATION OR BIOGENESIS
TISSUE DEVELOPMENT
DNA REPLICATION
STOMATAL COMPLEX DEVELOPMENT
DEVELOPMENTAL PROCESS
WATER TRANSPORT
CELL WALL ORGANIZATION OR BIOGENESIS
SULFUR COMPOUND METABOLIC PROCESS
MULTICELLULAR ORGANISMAL PROCESS
POSITIVE REGULATION OF ANTHOCYANIN BIOSYNTHETIC PROCESS
RESPONSE TO INSECT
CELL CYCLE
CELL DIVISION
REGULATION OF CELL CYCLE
CELL CYCLE PROCESS
MICROTUBULE-BASED PROCESS
MICROTUBULE-BASED MOVEMENT
NUCLEAR DIVISION
REGULATION OF CHROMOSOME SEGREGATION
log (p-value)

Figure S6

Fig. S6. **Gene expression patterns for the *sog1* DREM paths.** (**A**) Expression profiles of the individual genes in each path (S1-S5), where the paths shown in **Fig. 2A** are shown in grey. (**B**) Scaled Venn diagram showing the overlap of DE genes identified from the wild-type and *sog1* γ-IR time courses. (**C**) Expression profiles of the 1624 wt- and 218-*sog1*-specific-DE genes colored based on the *sog1* paths in which they reside. In all panels, the *y* axis indicates the $\log_2$ Fold Change in expression in response to γ-IR and the *x* axis indicates the time in minutes (') and/or hours (h). (**D**) Heatmap showing the enrichment of GO terms identified from DREM paths S1-S5 after REVIGO similarity filtering. Grey indicates a p-value >0.5. See also **Source Data 2**.

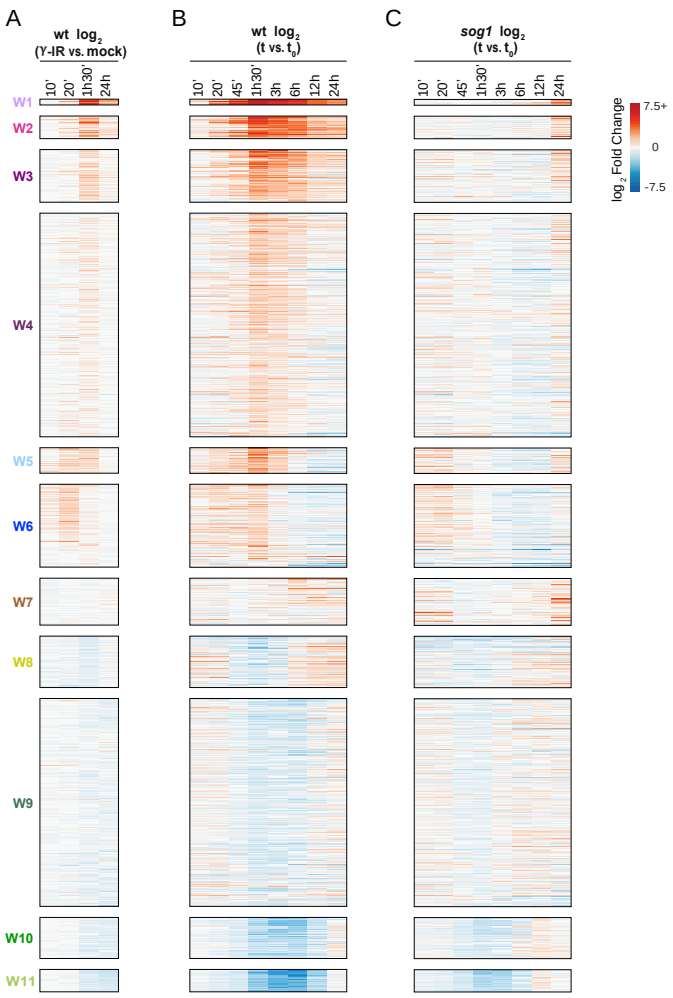

A
wt log2 (γ-IR vs. mock)
B
wt log2 (t vs. t0)
C
sog1 log2 (t vs. t0)
10'
20'
45'
1h30'
3h
6h
12h
24h
W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
W11
log2 Fold Change
7.5+
0
-7.5

Figure S7

Fig. S7. **Reproducible features of the wild-type and *sog1* γ-IR time courses.** (**A-C**) Heatmaps showing the $\log_2$ Fold Change in expression of the genes present in paths W1-W11 of the wild-type DREM model, ordered as in **Fig. 2C**. In **A**, the expression values represent γ-IR vs. mock treated samples at the indicated time points using wild-type seedlings, which constitutes an independent replicate of the experiment presented in **Fig. 1**, but over a more limited time course. In **B** and **C**, the expression values represent an alternate experimental design wherein the expression levels of γ-IR treated wild-type or *sog1* samples at each time point (t) were compared to the expression levels of un-irradiated samples at time zero (t_0). Although this t vs. t_0 setup was found to be less optimal, since changes in expression due to circadian rhythms and normal developmental processes were not well controlled for, it nonetheless serves as a useful control for features identified in the DREM model utilizing matched time points (γ-IR vs. mock). For comparison with **Fig. 2C**, the heatmaps were plotted on a $\log_2$ scale from -7.5 to +7.5.

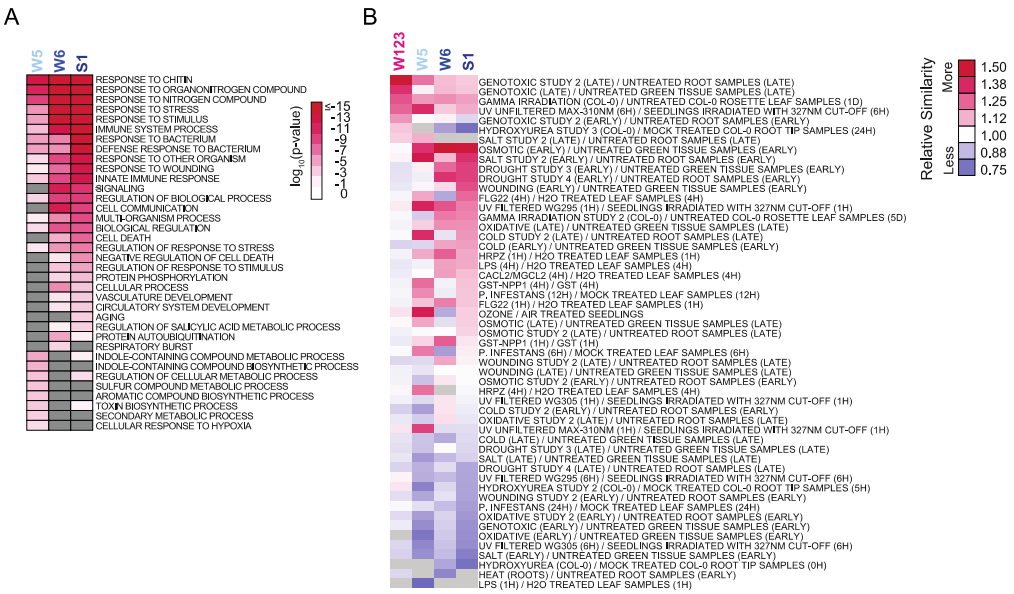

A
W5
W6
S1
RESPONSE TO CHITIN
RESPONSE TO ORGANONITROGEN COMPOUND
RESPONSE TO NITROGEN COMPOUND
RESPONSE TO STRESS
RESPONSE TO STIMULUS
IMMUNE SYSTEM PROCESS
RESPONSE TO BACTERIUM
DEFENSE RESPONSE TO BACTERIUM
RESPONSE TO OTHER ORGANISM
RESPONSE TO WOUNDING
INNATE IMMUNE RESPONSE
SIGNALING
REGULATION OF BIOLOGICAL PROCESS
CELL COMMUNICATION
MULTI-ORGANISM PROCESS
BIOLOGICAL REGULATION
CELL DEATH
REGULATION OF RESPONSE TO STRESS
NEGATIVE REGULATION OF CELL DEATH
REGULATION OF RESPONSE TO STIMULUS
PROTEIN PHOSPHORYLATION
CELLULAR PROCESS
VASCULATURE DEVELOPMENT
CIRCULATORY SYSTEM DEVELOPMENT
AGING
REGULATION OF SALICYLIC ACID METABOLIC PROCESS
PROTEIN AUTOUBIQUITINATION
RESPIRATORY BURST
INDOLE-CONTAINING COMPOUND METABOLIC PROCESS
INDOLE-CONTAINING COMPOUND BIOSYNTHETIC PROCESS
REGULATION OF CELLULAR METABOLIC PROCESS
SULFUR COMPOUND METABOLIC PROCESS
AROMATIC COMPOUND BIOSYNTHETIC PROCESS
TOXIN BIOSYNTHETIC PROCESS
SECONDARY METABOLIC PROCESS
CELLULAR RESPONSE TO HYPOXIA
log10 (p-value)
<-15
-13
-11
-9
-7
-5
-3
-1
0
B
W123
W5
W6
S1
GENOTOXIC STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
GENOTOXIC (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
GAMMA IRRADIATION (COL-0) / UNTREATED COL-0 ROSETTE LEAF SAMPLES (1D)
UV UNFILTERED MAX-310NM (6H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (6H)
GENOTOXIC STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
HYDROXYUREA STUDY 3 (COL-0) / MOCK TREATED COL-0 ROOT TIP SAMPLES (24H)
SALT STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
OSMOTIC (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
SALT STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
DROUGHT STUDY 3 (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
DROUGHT STUDY 4 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
WOUNDING (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
FLG22 (4H) / H2O TREATED LEAF SAMPLES (4H)
UV FILTERED WG295 (1H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (1H)
GAMMA IRRADIATION STUDY 2 (COL-0) / UNTREATED COL-0 ROSETTE LEAF SAMPLES (5D)
OXIDATIVE (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
COLD STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
COLD (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
HRPZ (1H) / H2O TREATED LEAF SAMPLES (1H)
LPS (4H) / H2O TREATED LEAF SAMPLES (4H)
CACL2/MGCL2 (4H) / H2O TREATED LEAF SAMPLES (4H)
GST-NPP1 (4H) / GST (4H)
P. INFESTANS (12H) / MOCK TREATED LEAF SAMPLES (12H)
FLG22 (1H) / H2O TREATED LEAF SAMPLES (1H)
OZONE / AIR TREATED SEEDLINGS
OSMOTIC (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
OSMOTIC STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
GST-NPP1 (1H) / GST (1H)
P. INFESTANS (6H) / MOCK TREATED LEAF SAMPLES (6H)
WOUNDING STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
WOUNDING (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
OSMOTIC STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
HRPZ (4H) / H2O TREATED LEAF SAMPLES (4H)
UV FILTERED WG305 (1H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (1H)
COLD STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
OXIDATIVE STUDY 2 (LATE) / UNTREATED ROOT SAMPLES (LATE)
UV UNFILTERED MAX-310NM (1H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (1H)
COLD (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
DROUGHT STUDY 3 (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
SALT (LATE) / UNTREATED GREEN TISSUE SAMPLES (LATE)
DROUGHT STUDY 4 (LATE) / UNTREATED ROOT SAMPLES (LATE)
UV FILTERED WG295 (6H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (6H)
HYDROXYUREA STUDY 2 (COL-0) / MOCK TREATED COL-0 ROOT TIP SAMPLES (5H)
WOUNDING STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
P. INFESTANS (24H) / MOCK TREATED LEAF SAMPLES (24H)
OXIDATIVE STUDY 2 (EARLY) / UNTREATED ROOT SAMPLES (EARLY)
GENOTOXIC (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
OXIDATIVE (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
UV FILTERED WG305 (6H) / SEEDLINGS IRRADIATED WITH 327NM CUT-OFF (6H)
SALT (EARLY) / UNTREATED GREEN TISSUE SAMPLES (EARLY)
HYDROXYUREA (COL-0) / MOCK TREATED COL-0 ROOT TIP SAMPLES (0H)
HEAT (ROOTS) / UNTREATED ROOT SAMPLES (EARLY)
LPS (1H) / H2O TREATED LEAF SAMPLES (1H)
Relative Similarity
More
Less
1.50
1.38
1.25
1.12
1.00
0.88
0.75

Figure S8

Fig. S8. **Analysis of the early, SOG1-independent transcriptional responses to DNA damage.** (**A**) Heatmap showing the enrichment of GO terms (log_{10} p-value < -1.7 in at least one path) for DREM paths W5, W6 and S1 after REVIGO similarity filtering. Grey indicates a p-value <0.5. See also **Source Data 2**. (**B**) Heatmap showing the degree of resemblance between the gene expression values of paths W1-W3 (wt; 3 h), W5 (wt; 1 h 30 min), W6 (wt; 20 min), S1 (*sog1*; 20 min) and the indicated stress datasets (See **supporting methods**) as determined using the Genevestigator signature tool (19). Grey values indicate categories not in the top 50 most similar datasets for a given path.


S1
5.1e-360; 277 (98%)
C2C2dof
5.0e-216; 284 (100%)
ABI3VP1
1.3e-100; 284 (100%)
BBRBPC
7.4e-37; 84 (30%)
AP2EREBP
1.9e-11; 284 (100%)
unknown
2.3e-5; 14 (39%)
unknown
4.6e-13; 45 (16%)
C2C2gata
S2
6.5e-586; 518 (100%)
C2C2dof
1.2e-204; 518 (100%)
C2H2
2.0e-169; 337 (65%)
REM
8.3e-23; 59 (11%)
AP2EREBP
8.9e-36; 518 (100%)
ABI3VP1
4.7e-4; 30 (6%)
unknown
1.1e-10; 18 (3%)
bHLH
S3
7.3e-839; 680 (100%)
C2C2dof
5.8e-317; 680 (100%)
C2H2
1.0e-288; 515 (76%)
ABI3VP1
1.8e-90; 378 (56%)
ABI3VP1
1.8e-61; 240 (35%)
FRS9
5.9e-43; 253 (37%)
ABI3VP1
2.0e-26; 116 (17%)
mTERF
3.7e-20; 74 (11%)
AP2EREBP
2.1e-10; 62 (9%)
unknown
9.2e-5; 62 (9%)
unknown
S4
5.0e-110; 828 (100%)
BBRBPC
5.4e-404; 830 (100%)
C2C2dof
6.0e-348; 553 (67%)
ABI3VP1
5.2e-144; 296 (36%)
FRS9
3.1e-120; 235 (28%)
AP2EREBP
4.4e-107; 441 (53%)
unknown
2.1e-59; 296 (36%)
MYBrelated
3.3e-38; 445 (54%)
ABI3VP1
1.3e-11; 110 (13%)
unknown
2.6e-8; 362 (44%)
unknown
S5
1.4e-77; 78 (95%)
MYB
2.9e-136; 82 (100%)
BBRBPC
3.0e-54; 81 (99%)
C2C2dof
8.0e-16; 81 (99%)
unknown
1.1e-15; 82 (100%)
ABI3VP1
3.4e-4; 49 (60%)
MYB
4.3e-17; 79 (96%)
BBRBPC
6.2e-17; 82 (100%)
unknown

Figure S9

Fig. S9. **Enriched motifs identified from the *sog1* DREM model.** Up to 10 motifs in the promoter regions (-500 to +350 bp around the TSS) of the genes in each DREM path were identified using the MEME tool (25) from the MEME suite (14). Below each motif, an E-value representing the significance of the motif is indicated along with the number and percentage of genes per path with the motif. In addition, the most similar TF match for each motif is indicated below each motif, where unknown = no match with an E-value <1e-2. Motifs that fulfill the following criteria are boxed and labeled in bold: First, they must be largely specific to a given path which excludes the BBRBPC, AP2EREBP, ABI3VP1, and C2C2dof motifs found in 3 or more of the 5 DREM paths. Second, they must not be low complexity, which excluded the REM motifs that consist in long stretches of A motifs. Finally, they must be present in >45% of the genes in a given path. See also **Source Data 3**.

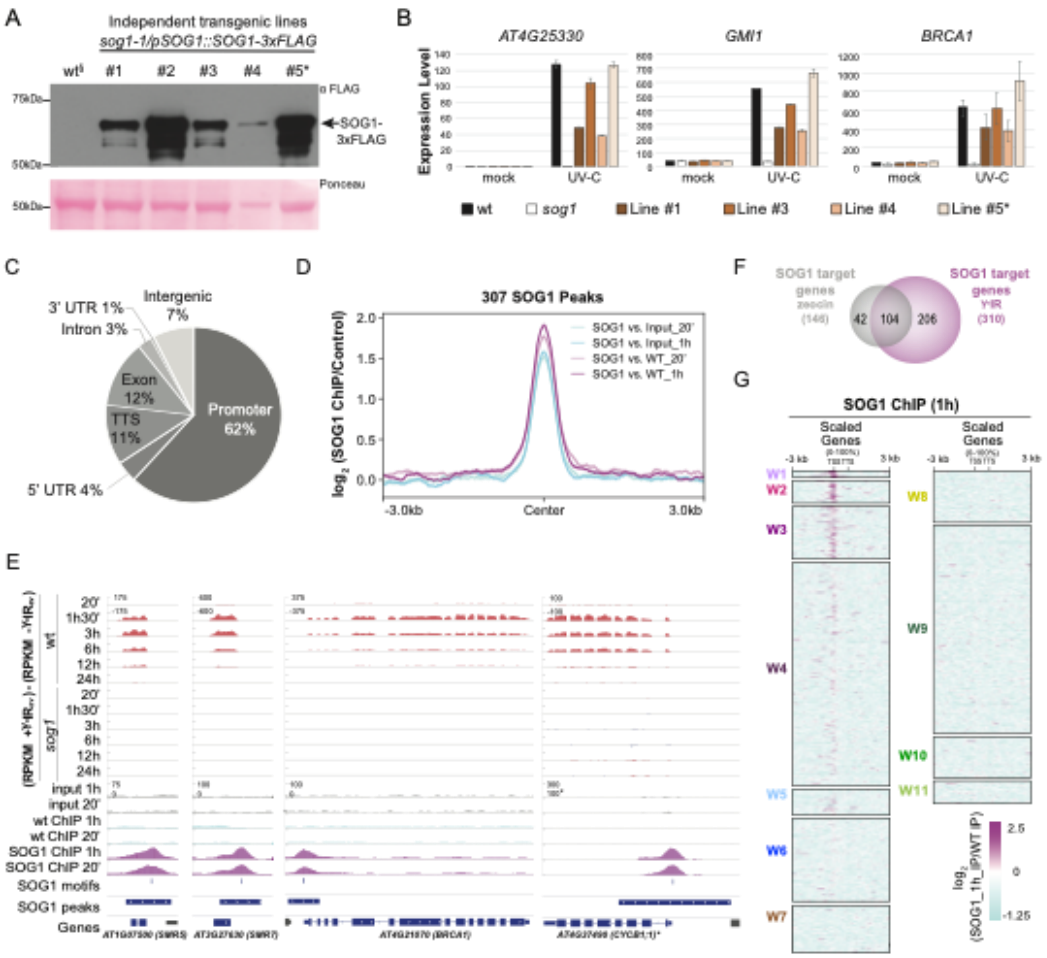

A
Independent transgenic lines
sog1-1/pSOG1::SOG1-3xFLAG
wt§ #1 #2 #3 #4 #5*
α FLAG
75kDa
50kDa
SOG1-3xFLAG
Ponceau
50kDa
B
AT4G25330
GMI1
BRCA1
Expression Level
mock
UV-C
wt
sog1
Line #1
Line #3
Line #4
Line #5*
C
3' UTR 1%
Intergenic 7%
Intron 3%
Exon 12%
TTS 11%
Promoter 62%
5' UTR 4%
D
307 SOG1 Peaks
log2 (SOG1 ChIP/Control)
SOG1 vs. Input_20'
SOG1 vs. Input_1h
SOG1 vs. WT_20'
SOG1 vs. WT_1h
-3.0kb
Center
3.0kb
E
(RPKM +YHR₀₀₇) - (RPKM +YHR₀₀₇)
wt
sog1
20'
1h30'
3h
6h
12h
24h
input 1h
input 20'
wt ChIP 1h
wt ChIP 20'
SOG1 ChIP 1h
SOG1 ChIP 20'
SOG1 motifs
SOG1 peaks
Genes
AT1G07500 (SWR5)
AT3G27630 (SWR1)
AT4G21070 (BRCA1)
AT4G37490 (CYCB1;1)*
F
SOG1 target genes zeocin (146)
42
104
206
SOG1 target genes YHR (310)
G
SOG1 ChIP (1h)
Scaled Genes
-3 kb
3 kb
W1
W2
W3
W4
W5
W6
W7
W8
W9
W10
W11
log2 (SOG1_1h_IP/WT IP)
2.5
0
-1.25

Figure S10

Fig. S10. ***sog1* complementation and SOG1 ChIP-seq analysis.** (**A**) Anti-FLAG western-blot detecting SOG1-3xFLAG protein levels 4 h after UV-C irradiation from five independent transgenic plant lines, as well as a control that lacks the SOG1 transgene (wt§). Ponceau staining of the membrane is shown below as a loading control. (**B**) qRT-PCR showing the relative expression levels of three genes (*AT4G25330, GMI1*, and *BRCA1*) induced by DNA damage in a SOG1-dependent manner (see **Dataset S4**) 3 h after either mock or UV-C treatments in the genetic backgrounds indicated below. For **A** and **B**, the SOG1-3xFLAG line marked with an asterisk (*) was used in the ChIP experiments. (**C**) Pie chart showing the fraction of SOG1 peaks that overlap with the indicated genome features. Here, the promoter regions are defined as -1 kb and +100 bp relative to the TSS. (**D**) Profile plot showing the enrichment levels of SOG1 at the peaks identified from both ChIP assays (20 min and 1 h) vs. their respective input and wt ChIP controls. (**E**) Screenshots showing the enrichment of SOG1 at select TAIR10 target genes relative to input and wt ChIP controls at both the 20 min and 1 h time points. Regions called as "SOG1 peaks" and the positions of "SOG1 motifs" are indicated below. The expression levels of these SOG1 target genes during both the wt and *sog1* γ-IR time courses are represented as the difference between RPKM normalized mock and γ-IR treated samples after averaging the data from two biological replicates. The scales for each set of tracks are indicated (upper left). At the *CYCB1;1* locus, the asterisk (*) indicates that the ChIP-seq tracks start at +100 because the *sog1* background harbors a transgenic version of the *CYCB1;1* promoter. (**F**) Scaled Venn diagrams showing the overlap of SOG1 target genes identified by ChIP-seq after zeocin (24) or γ-IR treatment. (**G**) Heatmaps, ordered as in **Fig. 2C**, showing the enrichment of SOG1 [$\log_2$ (SOG1/wt)] surrounding the TSS of the W1, W2, W3, and to a lesser extent, W4 genes from the wild-type DREM model.

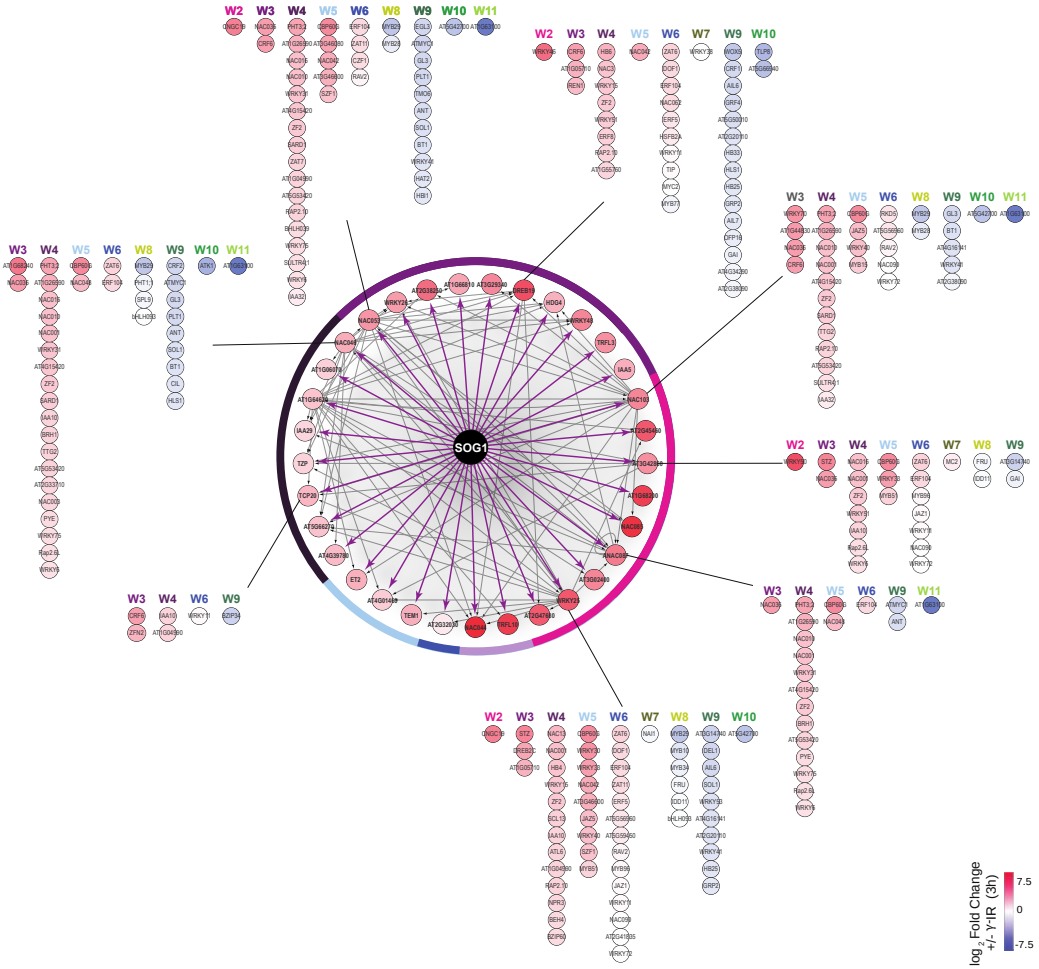


Figure S11

Fig. S11. **Transcription factor networks downstream of SOG1.** The 33 TF genes directly targeted by SOG1 (purple arrows) are displayed in the inner circle and are organized according to the DREM path they belong to, as indicated by the coloring of the outer, solid circle. Additional interactions amongst these direct SOG1 targets were added as grey arrows based on the DAP-seq database. A putative second layer of TF interactions was also added based on the targets of the 8 TFs present in the DAP-seq database (11). Each TF is colored according to the $\log_2$ Fold Change +/- γ-IR at the 3 h time point from the wild-type DREM model.

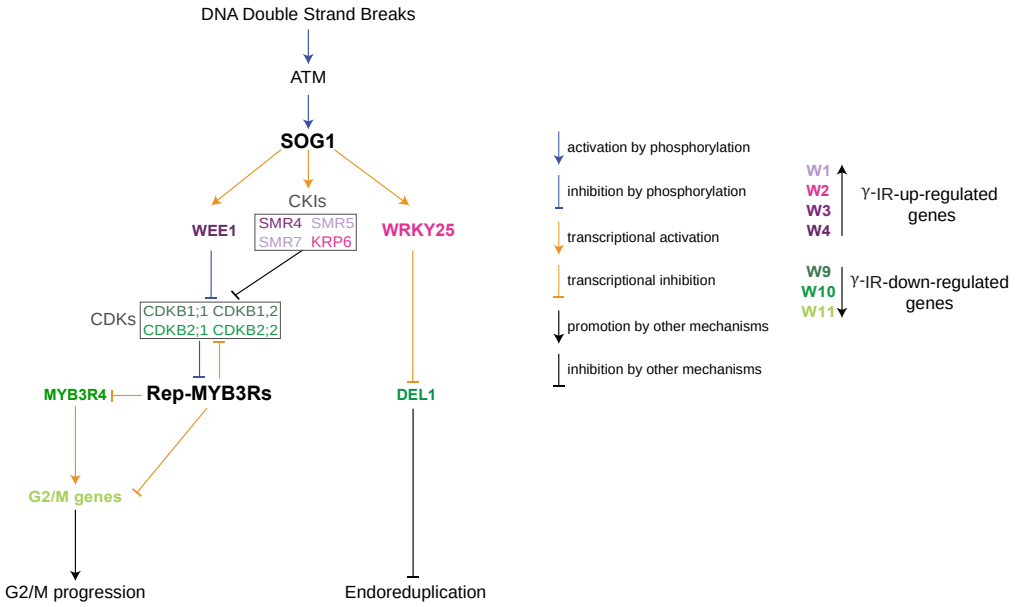

DNA Double Strand Breaks
ATM
SOG1
CKIs
WEE1
SMR4 SMR5
SMR7 KRP6
WRKY25
CDKs
CDKB1;1 CDKB1;2
CDKB2;1 CDKB2;2
MYB3R4
Rep-MYB3Rs
DEL1
G2/M genes
G2/M progression
Endoreduplication
activation by phosphorylation
inhibition by phosphorylation
transcriptional activation
transcriptional inhibition
promotion by other mechanisms
inhibition by other mechanisms
W1
W2
W3
W4
γ-IR-up-regulated genes
W9
W10
W11
γ-IR-down-regulated genes

Figure S12

Fig. S12. **Signaling cascades downstream of SOG1 controlling cell cycle arrest and endoreduplication**. In response to DNA DSBs, SOG1 is phosphorylated by ATM and binds to the promoters of several CDK inhibitor (CKI) genes, including *SMR4*, *SMR5*, *SMR7*, *KRP6*, and to the *WEE1* kinase gene, promoting their expression. These CKIs then inhibit the activity of CDKs, and this inhibition has been proposed to stabilize the Rep-MYB3R TFs, including MYB3R3 (35). MYB3R3, in turn, inhibits the expression of *CDKB2;1* and possibly other CDKs, creating a reinforcing loop. In addition, MYB3R3 represses many other G2/M genes (2), thereby contributing to the G2/M cell cycle arrest observed in response to DNA damage. This suppression is further enforced by the repression of *MYB3R4*, an activator MYB3R that regulates many G2/M genes (27). In addition to cell cycle arrest, there are also putative connections to endoreduplication. For example, *DEL1*, a negative regulator of endoreduplication (36), is bound by WRKY25 (11), a direct target of SOG1. Thus, it can be hypothesized that by activating *WRKY25* expression, SOG1 contributes to the repression of *DEL1* and the promotion of endoreduplication in response to DNA damage.

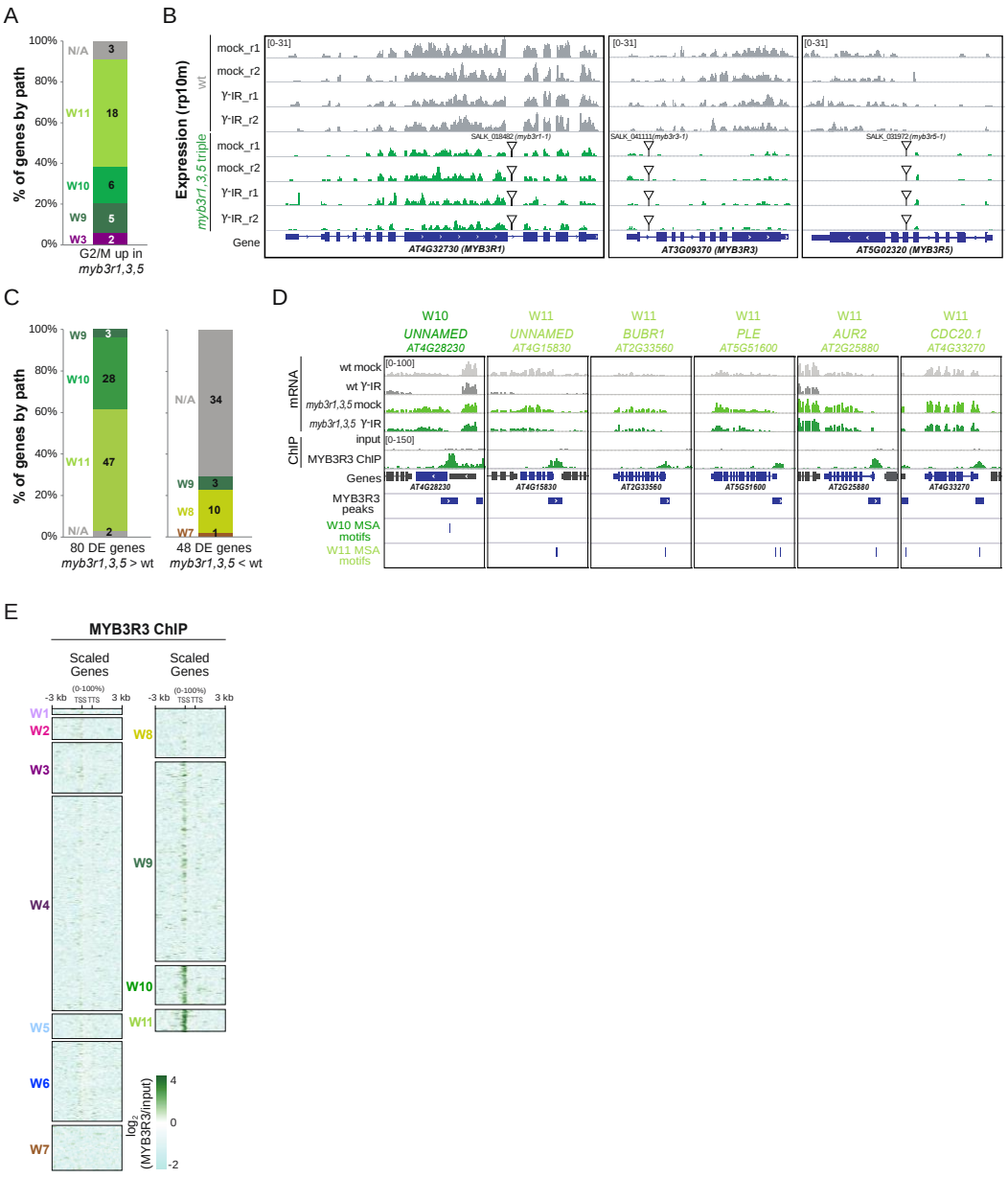

A
B
C
D
E
% of genes by path
G2/M up in myb3r1,3,5
Expression (rp10m)
wt
myb3r1,3,5 triple
mock_r1
mock_r2
γ-IR_r1
γ-IR_r2
Gene
AT4G32730 (MYB3R1)
AT3G09370 (MYB3R3)
AT5G02320 (MYB3R5)
80 DE genes myb3r1,3,5 > wt
48 DE genes myb3r1,3,5 < wt
mRNA
ChIP
wt mock
wt γ-IR
myb3r1,3,5 mock
myb3r1,3,5 γ-IR
input
MYB3R3 ChIP
Genes
MYB3R3 peaks
W10 MSA motifs
W11 MSA motifs
UNNAMED AT4G28230
UNNAMED AT4G15830
BUBR1 AT2G33560
PLE AT5G51600
AUR2 AT2G25880
CDC20.1 AT4G33270
MYB3R3 ChIP
Scaled Genes
log2 (MYB3R3/input)

Figure S13

Fig. S13. **Characterization of the Rep-MYBs within the context of the DREM model.** (**A**) Distribution of the G2/M genes less repressed in the *myb3r1,3,5* triple mutant (2) within the wild-type DREM paths. (**B**) Screenshots showing the expression levels of the three MYB3R genes in wild-type (wt) plants or the indicated mutant alleles. The scale is indicated in brackets. The gene orientations are indicated by the directions of the chevron symbols and the approximate locations of the T-DNA insertions for each allele are indicated by hollow black triangles. (**C**) Distribution of differentially expressed (DE) genes that are either less repressed (*myb3r1,3,5* > wt; Left) or less induced (*myb3r1,3,5* < wt; Right) relative to the wild-type controls after γ-IR. (**D**) Screenshots showing MYB3R3 peaks (2) at select TAIR10 target genes relative to input controls. Regions called as MYB3R3 peaks (2) and the positions of MSA motifs identified in the promoters of these genes using the MEME tool (25) from the MEME suite (14) (see **Source Data 3**) are indicated below. Expression levels of these MYB3R3 target genes in the wild-type or *myb3r1,3,5* triple mutant backgrounds after either mock or γ-IR treatments are shown above. The scales for each set of tracks are indicated in brackets. (**E**) Heatmaps, ordered as in **Fig. 2C**, showing the enrichment of MYB3R3 (2) [$\log_2$ (MYB3R3/input)] specifically at the TSS of 9, W10 and W11 genes from the wild-type DREM model.

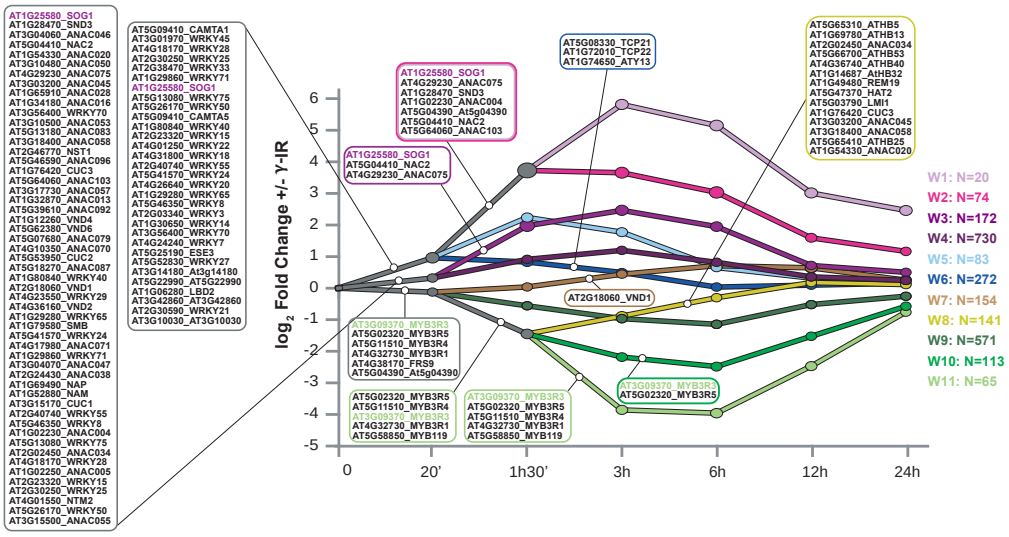

AT1G25580_SOG1
AT1G28470_SND3
AT3G04060_ANAC046
AT5G04410_NAC2
AT1G54330_ANAC020
AT3G10480_ANAC050
AT4G29230_ANAC075
AT3G03200_ANAC045
AT1G65910_ANAC028
AT1G34180_ANAC016
AT3G56400_WRKY70
AT3G10500_ANAC053
AT5G13180_ANAC083
AT3G18400_ANAC058
AT2G46770_NST1
AT5G46590_ANAC096
AT1G76420_CUC3
AT5G64060_ANAC103
AT3G17730_ANAC057
AT1G32870_ANAC013
AT5G39610_ANAC092
AT1G12260_VND4
AT5G62380_VND6
AT5G07680_ANAC079
AT4G10350_ANAC070
AT5G53950_CUC2
AT5G18270_ANAC087
AT1G80840_WRKY40
AT2G18060_VND1
AT4G23550_WRKY29
AT4G36160_VND2
AT1G29280_WRKY65
AT1G79580_SMB
AT5G41570_WRKY24
AT4G17980_ANAC071
AT1G29860_WRKY71
AT3G04070_ANAC047
AT2G24430_ANAC038
AT1G69490_NAP
AT1G52880_NAM
AT3G15170_CUC1
AT2G40740_WRKY55
AT5G46350_WRKY8
AT1G02230_ANAC004
AT5G13080_WRKY75
AT2G02450_ANAC034
AT4G18170_WRKY28
AT1G02250_ANAC005
AT2G23320_WRKY15
AT2G30250_WRKY25
AT4G01550_NTM2
AT5G26170_WRKY50
AT3G15500_ANAC055
AT5G09410_CAMTA1
AT3G01970_WRKY45
AT4G18170_WRKY28
AT2G30250_WRKY25
AT2G38470_WRKY33
AT1G29860_WRKY71
AT1G25580_SOG1
AT5G13080_WRKY75
AT5G26170_WRKY50
AT5G09410_CAMTA5
AT1G80840_WRKY40
AT2G23320_WRKY15
AT4G01250_WRKY22
AT4G31800_WRKY18
AT2G40740_WRKY55
AT5G41570_WRKY24
AT4G26640_WRKY20
AT1G29280_WRKY65
AT5G46350_WRKY8
AT2G03340_WRKY3
AT1G30650_WRKY14
AT3G56400_WRKY70
AT4G24240_WRKY7
AT5G25190_ESE3
AT5G52830_WRKY27
AT3G14180_At3g14180
AT5G22990_AT5G22990
AT1G06280_LBD2
AT3G42860_AT3G42860
AT2G30590_WRKY21
AT3G10030_AT3G10030
log2 Fold Change +/- γ-IR
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
0
20'
1h30'
3h
6h
12h
24h
AT1G25580_SOG1
AT4G29230_ANAC075
AT5G04390_At5g04390
AT1G02230_ANAC004
AT5G64060_ANAC103
AT1G25580_SOG1
AT5G04410_NAC2
AT4G29230_ANAC075
AT5G08330_TCP21
AT1G72010_TCP22
AT1G74650_ATY13
AT5G65310_ATHB5
AT1G69780_ATHB13
AT2G02450_ANAC034
AT5G66700_ATHB53
AT4G36740_ATHB40
AT1G14687_AtHB32
AT1G49480_REM19
AT5G47370_HAT2
AT5G03790_LMI1
AT1G76420_CUC3
AT3G03200_ANAC045
AT3G18400_ANAC058
AT5G65410_ATHB25
AT1G54330_ANAC020
AT2G18060_VND1
AT3G09370_MYB3R3
AT5G02320_MYB3R5
AT5G11510_MYB3R4
AT4G32730_MYB3R1
AT4G38170_FRS9
AT5G04390_At5g04390
AT5G02320_MYB3R5
AT5G11510_MYB3R4
AT3G09370_MYB3R3
AT4G32730_MYB3R1
AT5G58850_MYB119
AT3G09370_MYB3R3
AT5G02320_MYB3R5
AT5G02320_MYB3R5
AT5G11510_MYB3R4
AT4G32730_MYB3R1
AT5G58850_MYB119
W1: N=20
W2: N=74
W3: N=172
W4: N=730
W5: N=83
W6: N=272
W7: N=154
W8: N=141
W9: N=571
W10: N=113
W11: N=65

Figure S14

Fig. S14. **Final DREM model of the *Arabidopsis* DNA damage response.** The wild-type DREM model, as described in **SI Appendix, Fig. S4**, but including the ChIP-seq data for SOG1 and MYB3R3 (2). The new TF assignments for SOG1 and MYB3R3 are shown in purple and light green, respectively.

References

1. Yoshiyama K, Conklin PA, Huefner ND, & Britt AB (2009) Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A* 106(31):12843-12848.
2. Kobayashi K*, et al.* (2015) Transcriptional repression by MYB3R proteins regulates plant organ growth. *EMBO J* 34(15):1992-2007.
3. Dobin A*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.
4. Heinz S*, et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576-589.
5. Krishnakumar V*, et al.* (2015) Araport: the Arabidopsis information portal. *Nucleic Acids Res* 43(Database issue):D1003-1009.
6. Ramirez F*, et al.* (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44(W1):W160-165.
7. Kestler HA*, et al.* (2008) VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics* 9:67.
8. Davuluri RV*, et al.* (2003) AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinformatics* 4:25.
9. Palaniswamy SK*, et al.* (2006) AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiol* 140(3):818-829.
10. Yilmaz A*, et al.* (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Res* 39(Database issue):D1118-1122.
11. O'Malley RC*, et al.* (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 165(5):1280-1292.
12. Ernst J, Vainas O, Harbison CT, Simon I, & Bar-Joseph Z (2007) Reconstructing dynamic regulatory maps. *Mol Syst Biol* 3(1):74.
13. Schulz MH*, et al.* (2012) DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst Biol* 6(1):104.
14. Bailey TL*, et al.* (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue):W202-208.
15. Bailey TL, Boden M, Whitington T, & Machanick P (2010) The value of position-specific priors in motif discovery using MEME. *BMC Bioinformatics* 11:179.
16. Gupta S, Stamatoyannopoulos JA, Bailey TL, & Noble WS (2007) Quantifying similarity between motifs. *Genome Biol* 8(2):R24.
17. Boyle EI*, et al.* (2004) GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20(18):3710-3715.
18. Supek F, Bosnjak M, Skunca N, & Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.
19. Hruz T*, et al.* (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008:420747.
20. Law JA*, et al.* (2010) A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. *Curr Biol* 20(10):951-956.
21. Clough SJ & Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16(6):735-743.
22. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.

23. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
24. Ogita N*, et al.* (2018) Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in Arabidopsis. *Plant J*.
25. Bailey TL & Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28-36.
26. Menges M, de Jager SM, Gruissem W, & Murray JA (2005) Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J* 41(4):546-566.
27. Haga N*, et al.* (2011) Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in Arabidopsis. *Plant Physiol* 157(2):706-717.
28. Mi H*, et al.* (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45(D1):D183-D189.
29. Fischer M (2017) Census and evaluation of p53 target genes. *Oncogene* 36(28):3943-3956.
30. Kenzelmann Broz D*, et al.* (2013) Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. *Genes Dev* 27(9):1016-1031.
31. Wang Z, Schwacke R, & Kunze R (2016) DNA Damage-Induced Transcription of Transposable Elements and Long Non-coding RNAs in Arabidopsis Is Rare and ATM-Dependent. *Mol Plant* 9(8):1142-1155.
32. Culligan KM, Robertson CE, Foreman J, Doerner P, & Britt AB (2006) ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J* 48(6):947-961.
33. Bohmdorfer G*, et al.* (2011) GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in Arabidopsis. *Plant J* 67(3):420-433.
34. Ricaud L*, et al.* (2007) ATM-mediated transcriptional and developmental responses to gamma-rays in Arabidopsis. *PLoS One* 2(5):e430.
35. Chen P*, et al.* (2017) Arabidopsis R1R2R3-Myb proteins are essential for inhibiting cell division in response to DNA damage. *Nat Commun* 8(1):635.
36. Vlieghe K*, et al.* (2005) The DP-E2F-like gene DEL1 controls the endocycle in Arabidopsis thaliana. *Curr Biol* 15(1):59-63.

ACKNOWLEDGEMENTS

Chapter 2 is made up of a reprint of a published manuscript. The published manuscript is: Bourbousse, Clara; Vegesna, Neeraja and Law, Julie A. "**SOG1 activator and MYB3R repressors regulate a complex DNA damage network in *Arabidopsis***." Proceedings of the National Academy of Sciences 115.52 (2018): E12453-E12462. The dissertation author is the secondary author of this paper.

Chapter 3 EFFORTS TO IDENTIFY CHROMATIN- ASSOCIATED FACTORS INVOLVED IN DNA REPAIR

CHAPTER SUMMARY

Given the importance of various chromatin modifications in proper DNA repair, we hypothesized that both the pre-existing patterns of chromatin modifications at sites of DNA damage, and the changes in chromatin that occur during the repair process itself, affect the efficiency and fidelity of the DDR process. To understand how different chromatin modifications influence repair, it is important to identify and characterize the enzymes and pathways controlling the histone landscape and the recruitment of chromatin effectors to sites of DNA damage. Section **"A reverse genetic screen to identify chromatin-associated factors involved in DSB repair"** describes the pipeline we adopted for the identification of novel chromatin-associated factors required for DNA repair using a co-expression based reverse genetic screen. Section **"Two *Arabidopsis* YEATS domain proteins facilitate DNA damage repair via their histone binding domains"** describes the functional and mechanistic characterization of one candidate from the reverse genetic screen. Findings in this section are being prepared for submission for publication.

A REVERSE GENETIC SCREEN TO IDENTIFY CHROMATIN-ASSOCIATED FACTORS INVOLVED IN DSB REPAIR

ABSTRACT

DNA damage can arise from a multitude of endogenous and exogenous DNA damaging agents. To maintain genomic integrity, organisms detect DNA damage at chromatin and initiate the DNA Damage Response (DDR). Even though various chromatin modifications have been associated with repair, the roles of the factors interacting with them remains poorly understood. The following work describes the set-up and implementation of a coexpression-based reverse genetic screen to uncover key factors involved in linking chromatin and DNA repair in plants. Identification and characterization of such novel chromatin-associated factors and transcription factors will greatly enhance our knowledge about how chromatin modulates the DDR.

INTRODUCTION

In the nucleus, DNA is wrapped around histone proteins and is condensed into chromatin, which is the context in which DNA damage must be recognized and repaired. However, chromatin is not uniform. Both the DNA itself, as well as the histones, can be modified by the addition of post-translational modifications including acetylation, methylation, ubiquitination, and phosphorylation(Dona & Mittelsten Scheid, 2015; Vergara & Gutierrez, 2017). While many of these modifications have clear roles in regulating gene expression, current evidence suggests that they are also playing important roles in DNA repair. For example, the well characterized H2A.X phosphorylation(Friesner, Liu, Culligan, & Britt, 2005), and many other chromatin marks (discussed in detail in Chapter 1), have been associated with DNA repair. However, the roles of these modifications in orchestrating repair remains poorly understood. Moreover, the factors that

interact with these chromatin marks to facilitate DNA repair processes are yet to be identified and characterized.

To identify the roles of various chromatin modifications and associated chromatin effector proteins in repair, a model system, like *Arabidopsis*, where both the DNA repair machinery and chromatin modifications can me modulated genetically (*i.e.* mutants are not lethal) is essential. Furthermore, high levels of conservation in both DNA repair and chromatin modifying pathways among eukaryotes, the small size of the *Arabidopsis* genome, its rapid life cycle, high fecundity, and the availability of gene manipulation techniques make it an attractive model for DNA repair studies.

Several key players involved in DNA repair pathways in *Arabidopsis* were identified by forward genetic screens and *in silico* approaches. The initial genetic screens assessed fertility defects and genotoxic sensitivity by scoring morphological phenotypes arising due to defects in cell growth. The aforementioned phenotypes were specifically screened for, since improper meiosis, which involves resolution of programmed DNA double strand breaks, can result in fertility defects and DNA damage causing cell cycle arrest or in severe cases, cell death in the meristematic tissue, leads to abnormal growth of true leaves and roots. Although several key players in the repair pathways like *ULTRAVIOLET HYPERSENSITIVE 1 (UVH1)*(Harlow, Jenkins, Pittalwala, & Mount, 1994), *UV REPAIR DEFECTIVE 1 (UVR1)*(Britt, Chen, Wykoff, & Mitchell, 1993), *SUPPRESSOR OF GAMMA RADIATION 1 (SOG1)*(Preuss & Britt, 2003), etc. have been identified in these studies, these screens are biased towards identifying genes in cell growth associated processes. Moreover, they failed to identify many conserved DDR machinery characterized in yeast and mammals. Instead, homologs of the conserved DDR machinery were identified in plants using *in silico* approaches and their roles in repair were characterized using

molecular repair assays, like the homologous recombination (HR) reporter assays, instead of growth assays(Block-Schmidt, Dukowic-Schulze, Wanieck, Reidt, & Puchta, 2011; Hartung, Suer, & Puchta, 2007; Orel, Kyryk, & Puchta, 2003; Roth et al., 2012). Given the strengths of these molecular assays in identifying genes involved in DNA repair, several reporter assays were developed and utilized for screening. For example, a luciferase based system to monitor somatic HR events was used to identify homologous recombination associated proteins INO80(Fritsch, Benvenuto, Bowler, Molinier, & Hohn, 2004) and Polδ(Schuermann, Fritsch, Lucht, & Hohn, 2009). However, as the read-out of this assay requires both proper repair and proper expression of the transgenic reporter, both repair mutants and gene expression mutants would have been recovered, confounding the interpretation of these types of screens(Puchta & Hohn, 2012; Ulker et al., 2012).

Based on the mutants identified in the previous studies from the aforementioned screens and *in silico* studies, our understanding of how chromatin influences DNA repair still remains obscure, suggesting that many of the proteins and pathways regulating this process are yet to be identified. To identify novel chromatin-associated factors and avoid the drawbacks of the previously used systems, a reverse genetics approach based on gene co-expression after DNA damage was implemented. In short, it involves (1) selection of candidate genes based on their co-expression with known DNA repair factors, (2) procurement of mutant alleles and (3) implementing a multipronged approach to screen for one or more morphological and molecular DNA repair phenotypes. This reverse genetic approach to identify novel chromatin factors is advantageous over the previously discussed strategies, as it narrows down the candidate list to a relatively small set of candidate genes (~50), and thus, enables the use of multiple assays including, the labor-intensive, but higher resolution, molecular DNA repair assays. Moreover, it enables the

identification of DDR genes that show minimal or no phenotypes due to redundancy (as strategic higher order mutants of related factors can be generated and tested) and unlike *in silico* approaches, it should be able to recover plant specific factors.

SCREENING METHOD

Mining for candidate chromatin-associated proteins involved in DNA damage repair.

The 'Guilt by Association' principle states that genes with related function are often co-expressed, and in some cases form protein complexes(Wolfe, Kohane, & Butte, 2005). Leveraging this principle, the previously generated DDR co-expression network (Chapter 2, DREM network)(Bourbousse, Vegesna, & Law, 2018) was used to identify genes co-expressed with known DNA repair machinery to serve as a pool of potential new candidate chromatin factors involved in the repair process.

Three of the upregulated paths (W1, W2, W3) in the wild-type DREM model (**Figure 2-1A**), are enriched for DNA damage repair-associated genes with high confidence(Bourbousse et al., 2018). Among the 266 genes belonging to these three paths, 22 have characterized roles in DNA repair(Aklilu, Soderquist, & Culligan, 2014; Deveaux, Alonso, Pierrugues, Godon, & Kazmaier, 2000; Dong, Cai, & Makaroff, 2001; Garcia-Ortiz, Ariza, Hoffman, Hays, & Roldan-Arjona, 2004; Hartung, Plchova, & Puchta, 2000; Hirakawa, Hasegawa, White, & Matsunaga, 2017; Islam, Hosen, Zaman, & Islam, 2013; Jenkins et al., 1995; Jia, den Dulk-Ras, Shen, Hooykaas, & de Pater, 2013; Klutstein et al., 2008; Lafarge & Montane, 2003; Panoli et al., 2006; Pedroza-Garcia, Najera-Martinez, de la Paz Sanchez, & Plasencia, 2015; Sebastian et al., 2009; Shultz, Tatineni, Hanley-Bowdoin, & Thompson, 2007; C. Wang & Liu, 2006; Watanabe et al., 2009; Yin et al., 2009) and 5 are known to regulate chromatin organization(Cao & Jacobsen, 2002;

Sebastian et al., 2009; Woo, Pontes, Pikaard, & Richards, 2007; Zacharaki et al., 2012). The remaining uncharacterized genes were used to generate a list of candidate chromatin-associated factors with roles in DNA damage repair. Their putative roles were determined using PANTHER (Protein ANalysis THrough Evolutionary Relationships)(Mi et al., 2017), which identifies genes that are homologs of known chromatin effectors and DNA repair factors in different species and; SMART (Simple Modular Architecture Research Tool)(Letunic, Doerks, & Bork, 2015), which identifies genes with known chromatin-associated domains, including DNA binding domains (*e.g.* Zinc finger, MYB domains, etc.) and histone binding domains (*e.g.* YEATS, SAWADEE, etc.). Using this strategy, 53 candidate genes, including 4 non-coding RNAs, 9 putative repair factors, 11 putative chromatin-associated factors, 29 transcription factors and coactivators, were selected. In addition to these genes, close homologs of candidate genes that might function redundantly were also added to the screening pool (for higher order mutant generation). Overall, this selection approach narrowed the list of candidate genes from 266 to 53—an implementable number of candidates to screen using multiple repair assays.

Procurement of mutant alleles for candidate DNA repair factors.

As a first step towards investigating the functions of the candidate genes using a genetic approach, mutants were procured from several large collections of sequence-indexed T-DNA insertion mutants that are available for *Arabidopsis thaliana* (Alonso et al., 2003; Kleinboelting, Huep, Kloetgen, Viehoever, & Weisshaar, 2012; Li, Rosso, Strizhov, Viehoever, & Weisshaar, 2003; Samson et al., 2002; Sessions et al., 2002; Woody, Austin-Phillips, Amasino, & Krysan, 2007). Whenever possible, at least two mutant alleles per gene were selected to serve as biological replicates and to increase confidence in the assignment of the repair defects to the gene of interest.

In addition, considering that mutants defective in repair are prone to genomic instability and/or meiotic defects, homozygous mutant lines along with co-segregating heterozygous (in case of lethality in homozygous mutant) and wild-type plants (to act as controls) were generated. Two independent T-DNA mutant alleles for 30% of candidate genes and one mutant allele for another 35% of candidate genes were procured from available databases. For the remaining 35% of the candidate genes that lack available T-DNA insertional mutant lines, the CRISPR/Cas9 system is being used to generate homozygous mutant lines using three independent guide RNAs(Z. P. Wang et al., 2015; Xie, Zhang, & Yang, 2014), prioritizing putative repair factors and chromatin modifiers. While the generation of CRISPR alleles is still ongoing, the procured T-DNA mutants were made homozygous and subjected to one or more DNA repair assays described below.

Screening using DNA damage assays.

Defects in the DDR can be quantified using different DNA damage sensitivity assays. These assays include high-throughput growth assays (meiotic defects (De Muyt et al., 2009), true leaf assay(Rosa & Scheid, 2014a) and root length assay(Slovak et al., 2014)), which rely on plant growth phenotypes, and medium-throughput molecular assays (recombination GUS reporter assays(Puchta & Hohn, 2012; Roth et al., 2012) and comet assays(Angelis, Dušinská, & Collins, 1999; Hasplova et al., 2012; Kushwaha, Vikram, Trivedi, & Jena, 2011; Nagarajan et al., 2017)), which measure DNA repair more directly. In these assays, the defects in DNA damage can be assessed in response to a wide range of DNA damaging agents including: γ-irradiation and bleomycin, which produce DSB breaks; mitomycin C, a DNA intrastrand cross-linker; hydroxyurea, an inhibitor of DNA synthesis and UV-C, which results in photoproducts. Assessing

DNA repair defects in response to different DNA damaging agents allows identification of the potential repair pathways affected in a particular mutant.

Meiotic defects are expected in some candidate gene mutants. Meiosis and somatic DSB repair via homologous recombination both involve the formation and resolution of double strand breaks and thus, both processes employ some common machinery(De Muyt et al., 2007; Mercier, Mezard, Jenczewski, Macaisne, & Grelon, 2015). To screen for meiotic defects in candidate gene mutants, fertility was assessed by the number of seeds produced, pollen viability was assessed using alexander staining(Alexander, 1969; Peterson, Slovin, & Chen, 2010) and chromosomal segregation defects were assessed using meiotic chromosomal spreads and Fluorescence In Situ Hybridization (FISH)(Higgins, Wright, Bomblies, & Franklin, 2014). Moreover, if any of the candidates has a more rudimentary and intractable role in repair, they might not be viable and show embryonic defects(C. Wang & Liu, 2006), which can be assessed by aborted seeds in siliques and inability to recover homozygous mutants.

Growth defects are expected in candidate gene mutants if the candidate is involved in cell cycle arrest or programmed cell death during DNA damage. Sensitivity of mutants to different genotoxic agents was assessed based on easily determined phenotypes such as proper development of true leaves (true leaf assay)(Rosa & Scheid, 2014b) and a high throughput quantitative measurement of root lengths (root length estimation using Busch-lab Root Analysis Toolchain-BRAT) (Slovak et al., 2014). Their roles in cell growth or death-associated processes can be further assessed using cell cycle reporters(Culligan, Tissier, & Britt, 2004) and propidium iodide staining to check for cell viability (Takahashi et al., 2019).

To assess somatic homologous recombination repair defects in candidate genes, transgenic homologous recombination (HR) reporter assays developed and characterized by other

groups(Puchta & Hohn, 2012; Roth et al., 2012) were used. This system involves a transgenic, non-functional *β-glucuronidase (GUS)* gene, which upon DSB and repair via HR-like pathways restores the proper GUS sequence. Histochemical GUS staining of these plants provides a visible and quantitative readout of DSB repair events in the form of blue sectors on the seedling. To assess HR-like roles in candidate genes, DSB repair rates (indicated by number of blue sectors/seedling) were scored in mutants by introducing these well characterized reporter lines into mutant backgrounds by crossing.

To assess DNA repair defects in candidate genes directly at the molecular level without the aid of transgenic reporters (which require labor intensive genetics), the comet assay(Angelis et al., 1999; Hasplova et al., 2012; Kushwaha et al., 2011; Nagarajan et al., 2017) was used. The comet assay involves electrophoresis of agarose embedded nuclei extracted from DNA damaged tissue. Electrophoresis results in structures resembling comets when observed by fluorescence microscopy.

CONCLUSIONS

Using this reverse genetic screening approach, five candidate genes (and their homologs) have been identified and their characterization is currently ongoing. Some notable candidates that have shown promising DNA repair phenotypes include (1) histone reader proteins, YAF9B and its homolog YAF9A, which have been functionally and mechanistically characterized in the next section of this thesis and (2) the DNA repair protein, NSE4A, which has recently shown to be involved in DNA damage repair and seed development(Diaz et al., 2019). The characterization of these candidate genes, which revealed key roles in chromatin-associated aspects of the DNA damage response, validates the set-up of the screen to identify such factors.

REFERENCES

Aklilu, B. B., Soderquist, R. S., & Culligan, K. M. (2014). Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. *Nucleic Acids Res, 42*(5), 3104-3118. doi:10.1093/nar/gkt1292

Alexander, M. P. (1969). Differential staining of aborted and nonaborted pollen. *Stain Technol, 44*(3), 117-122. doi:10.3109/10520296909063335

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., . . . Ecker, J. R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science, 301*(5633), 653-657. doi:10.1126/science.1086391

Angelis, K. J., Dušinská, M., & Collins, A. R. (1999). Single cell gel electrophoresis: Detection of DNA damage at different levels of sensitivity. *Electrophoresis, 20*(10), 2133-2138. doi:10.1002/(sici)1522-2683(19990701)20:10<2133::Aid-elps2133>3.0.Co;2-q

Block-Schmidt, A. S., Dukowic-Schulze, S., Wanieck, K., Reidt, W., & Puchta, H. (2011). BRCC36A is epistatic to BRCA1 in DNA crosslink repair and homologous recombination in Arabidopsis thaliana. *Nucleic Acids Res, 39*(1), 146-154. doi:10.1093/nar/gkq722

Bourbousse, C., Vegesna, N., & Law, J. A. (2018). SOG1 activator and MYB3R repressors regulate a complex DNA damage network in Arabidopsis. *Proc Natl Acad Sci U S A, 115*(52), E12453-E12462. doi:10.1073/pnas.1810582115

Britt, A. B., Chen, J. J., Wykoff, D., & Mitchell, D. (1993). A UV-sensitive mutant of Arabidopsis defective in the repair of pyrimidine-pyrimidinone(6-4) dimers. *Science, 261*(5128), 1571-1574. doi:10.1126/science.8372351

Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci U S A, 99 Suppl 4*(suppl 4), 16491-16498. doi:10.1073/pnas.162371599

Culligan, K., Tissier, A., & Britt, A. (2004). ATR regulates a G2-phase cell-cycle checkpoint in Arabidopsis thaliana. *Plant Cell, 16*(5), 1091-1104. doi:10.1105/tpc.018903

De Muyt, A., Pereira, L., Vezon, D., Chelysheva, L., Gendrot, G., Chambon, A., . . . Grelon, M. (2009). A high throughput genetic screen identifies new early meiotic recombination functions in Arabidopsis thaliana. *PLoS Genet, 5*(9), e1000654. doi:10.1371/journal.pgen.1000654

De Muyt, A., Vezon, D., Gendrot, G., Gallois, J. L., Stevens, R., & Grelon, M. (2007). AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana. *EMBO J, 26*(18), 4126-4137. doi:10.1038/sj.emboj.7601815

Deveaux, Y., Alonso, B., Pierrugues, O., Godon, C., & Kazmaier, M. (2000). Molecular cloning and developmental expression of AtGR1, a new growth-related Arabidopsis gene strongly

induced by ionizing radiation. *Radiat Res, 154*(4), 355-364. doi:10.1667/0033-7587(2000)154[0355:mcadeo]2.0.co;2

Diaz, M., Pecinkova, P., Nowicka, A., Baroux, C., Sakamoto, T., Gandha, P. Y., . . . Pecinka, A. (2019). The SMC5/6 Complex Subunit NSE4A Is Involved in DNA Damage Repair and Seed Development. *Plant Cell, 31*(7), 1579-1597. doi:10.1105/tpc.18.00043

Dona, M., & Mittelsten Scheid, O. (2015). DNA Damage Repair in the Context of Plant Chromatin. *Plant Physiol, 168*(4), 1206-1218. doi:10.1104/pp.15.00538

Dong, F., Cai, X., & Makaroff, C. A. (2001). Cloning and characterization of two Arabidopsis genes that belong to the RAD21/REC8 family of chromosome cohesin proteins. *Gene, 271*(1), 99-108. doi:10.1016/s0378-1119(01)00499-1

Friesner, J. D., Liu, B., Culligan, K., & Britt, A. B. (2005). Ionizing radiation-dependent gamma-H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. *Mol Biol Cell, 16*(5), 2566-2576. doi:10.1091/mbc.e04-10-0890

Fritsch, O., Benvenuto, G., Bowler, C., Molinier, J., & Hohn, B. (2004). The INO80 protein controls homologous recombination in Arabidopsis thaliana. *Mol Cell, 16*(3), 479-485. doi:10.1016/j.molcel.2004.09.034

Garcia-Ortiz, M. V., Ariza, R. R., Hoffman, P. D., Hays, J. B., & Roldan-Arjona, T. (2004). Arabidopsis thaliana AtPOLK encodes a DinB-like DNA polymerase that extends mispaired primer termini and is highly expressed in a variety of tissues. *Plant J, 39*(1), 84-97. doi:10.1111/j.1365-313X.2004.02112.x

Harlow, G. R., Jenkins, M. E., Pittalwala, T. S., & Mount, D. W. (1994). Isolation of uvh1, an Arabidopsis mutant hypersensitive to ultraviolet light and ionizing radiation. *Plant Cell, 6*(2), 227-235. doi:10.1105/tpc.6.2.227

Hartung, F., Plchova, H., & Puchta, H. (2000). Molecular characterisation of RecQ homologues in Arabidopsis thaliana. *Nucleic Acids Res, 28*(21), 4275-4282. doi:10.1093/nar/28.21.4275

Hartung, F., Suer, S., & Puchta, H. (2007). Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in Arabidopsis thaliana. *Proc Natl Acad Sci U S A, 104*(47), 18836-18841. doi:10.1073/pnas.0705998104

Hasplova, K., Hudecova, A., Magdolenova, Z., Bjoras, M., Galova, E., Miadokova, E., & Dusinska, M. (2012). DNA alkylation lesions and their repair in human cells: modification of the comet assay with 3-methyladenine DNA glycosylase (AlkD). *Toxicol Lett, 208*(1), 76-81. doi:10.1016/j.toxlet.2011.10.005

Higgins, J. D., Wright, K. M., Bomblies, K., & Franklin, F. C. (2014). Cytological techniques to analyze meiosis in Arabidopsis arenosa for investigating adaptation to polyploidy. *Front Plant Sci, 4*, 546. doi:10.3389/fpls.2013.00546

Hirakawa, T., Hasegawa, J., White, C. I., & Matsunaga, S. (2017). RAD54 forms DNA repair foci in response to DNA damage in living plant cells. *Plant J, 90*(2), 372-382. doi:10.1111/tpj.13499

Islam, M. R., Hosen, M. I., Zaman, A., & Islam, M. O. (2013). Structural, functional and molecular docking study to characterize GMI1 from Arabidopsis thaliana. *Interdiscip Sci, 5*(1), 13-22. doi:10.1007/s12539-013-0153-1

Jenkins, M. E., Harlow, G. R., Liu, Z., Shotwell, M. A., Ma, J., & Mount, D. W. (1995). Radiation-sensitive mutants of Arabidopsis thaliana. *Genetics, 140*(2), 725-732.

Jia, Q., den Dulk-Ras, A., Shen, H., Hooykaas, P. J., & de Pater, S. (2013). Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in Arabidopsis thaliana. *Plant Mol Biol, 82*(4-5), 339-351. doi:10.1007/s11103-013-0065-9

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., & Weisshaar, B. (2012). GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res, 40*(Database issue), D1211-1215. doi:10.1093/nar/gkr1047

Klutstein, M., Shaked, H., Sherman, A., Avivi-Ragolsky, N., Shema, E., Zenvirth, D., . . . Simchen, G. (2008). Functional conservation of the yeast and Arabidopsis RAD54-like genes. *Genetics, 178*(4), 2389-2397. doi:10.1534/genetics.108.086777

Kushwaha, S., Vikram, A., Trivedi, P. P., & Jena, G. B. (2011). Alkaline, Endo III and FPG modified comet assay as biomarkers for the detection of oxidative DNA damage in rats with experimentally induced diabetes. *Mutat Res, 726*(2), 242-250. doi:10.1016/j.mrgentox.2011.10.004

Lafarge, S., & Montane, M. H. (2003). Characterization of Arabidopsis thaliana ortholog of the human breast cancer susceptibility gene 1: AtBRCA1, strongly induced by gamma rays. *Nucleic Acids Res, 31*(4), 1148-1155. doi:10.1093/nar/gkg202

Letunic, I., Doerks, T., & Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res, 43*(Database issue), D257-260. doi:10.1093/nar/gku949

Li, Y., Rosso, M. G., Strizhov, N., Viehoever, P., & Weisshaar, B. (2003). GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana. *Bioinformatics, 19*(11), 1441-1442. doi:10.1093/bioinformatics/btg170

Mercier, R., Mezard, C., Jenczewski, E., Macaisne, N., & Grelon, M. (2015). The molecular biology of meiosis in plants. *Annu Rev Plant Biol, 66*, 297-327. doi:10.1146/annurev-arplant-050213-035923

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome

pathways, and data analysis tool enhancements. *Nucleic Acids Res, 45*(D1), D183-D189. doi:10.1093/nar/gkw1138

Nagarajan, R., Muthusamy, G., Balupillai, A., Govindasamy, K., Ramasamy, K., Ponniresan, V., & Malla, I. (2017). Modified comet assays for the detection of cyclobutane pyrimidine dimers and oxidative base damages. *Journal of Radiation and Cancer Research, 8*(1), 82--86. doi:10.4103/0973-0168.199312

Orel, N., Kyryk, A., & Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J, 35*(5), 604-612. doi:10.1046/j.1365-313x.2003.01832.x

Panoli, A. P., Ravi, M., Sebastian, J., Nishal, B., Reddy, T. V., Marimuthu, M. P., . . . Siddiqi, I. (2006). AtMND1 is required for homologous pairing during meiosis in Arabidopsis. *BMC Mol Biol, 7*(1), 24. doi:10.1186/1471-2199-7-24

Pedroza-Garcia, J. A., Najera-Martinez, M., de la Paz Sanchez, M., & Plasencia, J. (2015). Arabidopsis thaliana thymidine kinase 1a is ubiquitously expressed during development and contributes to confer tolerance to genotoxic stress. *Plant Mol Biol, 87*(3), 303-315. doi:10.1007/s11103-014-0277-7

Peterson, R., Slovin, J. P., & Chen, C. (2010). A simplified method for differential staining of aborted and non-aborted pollen grains. *International Journal of Plant Biology, 1*(2), 13. doi:10.4081/pb.2010.e13

Preuss, S. B., & Britt, A. B. (2003). A DNA-damage-induced cell cycle checkpoint in Arabidopsis. *Genetics, 164*(1), 323-334.

Puchta, H., & Hohn, B. (2012). In planta somatic homologous recombination assay revisited: a successful and versatile, but delicate tool. *Plant Cell, 24*(11), 4324-4331. doi:10.1105/tpc.112.101824

Rosa, M., & Scheid, O. M. (2014a). DNA Damage Sensitivity Assays with Arabidopsis Seedlings. *Bio Protoc, 4*(7). doi:10.21769/BioProtoc.1093

Rosa, M., & Scheid, O. M. (2014b). Measuring Homologous Recombination Frequency in Arabidopsis Seedlings. *Bio Protoc, 4*(7). doi:10.21769/BioProtoc.1094

Roth, N., Klimesch, J., Dukowic-Schulze, S., Pacher, M., Mannuss, A., & Puchta, H. (2012). The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. *Plant J, 72*(5), 781-790. doi:10.1111/j.1365-313X.2012.05119.x

Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., . . . Lecharny, A. (2002). FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of Arabidopsis thaliana T-DNA transformants. *Nucleic Acids Res, 30*(1), 94-97. doi:10.1093/nar/30.1.94

Schuermann, D., Fritsch, O., Lucht, J. M., & Hohn, B. (2009). Replication stress leads to genome instabilities in Arabidopsis DNA polymerase delta mutants. *Plant Cell, 21*(9), 2700-2714. doi:10.1105/tpc.109.069682

Sebastian, J., Ravi, M., Andreuzza, S., Panoli, A. P., Marimuthu, M. P., & Siddiqi, I. (2009). The plant adherin AtSCC2 is required for embryogenesis and sister-chromatid cohesion during meiosis in Arabidopsis. *Plant J, 59*(1), 1-13. doi:10.1111/j.1365-313X.2009.03845.x

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., . . . Goff, S. A. (2002). A high-throughput Arabidopsis reverse genetics system. *Plant Cell, 14*(12), 2985-2994. doi:10.1105/tpc.004630

Shultz, R. W., Tatineni, V. M., Hanley-Bowdoin, L., & Thompson, W. F. (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. *Plant Physiol, 144*(4), 1697-1714. doi:10.1104/pp.107.101105

Slovak, R., Goschl, C., Su, X., Shimotani, K., Shiina, T., & Busch, W. (2014). A Scalable Open-Source Pipeline for Large-Scale Root Phenotyping of Arabidopsis. *Plant Cell, 26*(6), 2390-2403. doi:10.1105/tpc.114.124032

Takahashi, N., Ogita, N., Takahashi, T., Taniguchi, S., Tanaka, M., Seki, M., & Umeda, M. (2019). A regulatory module controlling stress-induced cell cycle arrest in Arabidopsis. *Elife, 8*. doi:10.7554/eLife.43944

Ulker, B., Hommelsheim, C. M., Berson, T., Thomas, S., Chandrasekar, B., Olcay, A. C., . . . Frantzeskakis, L. (2012). Reevaluation of the reliability and usefulness of the somatic homologous recombination reporter lines. *Plant Cell, 24*(11), 4314-4323. doi:10.1105/tpc.112.100404

Vergara, Z., & Gutierrez, C. (2017). Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome Biol, 18*(1), 96. doi:10.1186/s13059-017-1236-9

Wang, C., & Liu, Z. (2006). Arabidopsis ribonucleotide reductases are critical for cell cycle progression, DNA damage repair, and plant development. *Plant Cell, 18*(2), 350-365. doi:10.1105/tpc.105.037044

Wang, Z. P., Xing, H. L., Dong, L., Zhang, H. Y., Han, C. Y., Wang, X. C., & Chen, Q. J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol, 16*, 144. doi:10.1186/s13059-015-0715-0

Watanabe, K., Pacher, M., Dukowic, S., Schubert, V., Puchta, H., & Schubert, I. (2009). The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in Arabidopsis thaliana. *Plant Cell, 21*(9), 2688-2699. doi:10.1105/tpc.108.060525

Wolfe, C. J., Kohane, I. S., & Butte, A. J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *Bmc Bioinformatics, 6*, 227. doi:10.1186/1471-2105-6-227

Woo, H. R., Pontes, O., Pikaard, C. S., & Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev, 21*(3), 267-277. doi:10.1101/gad.1512007

Woody, S. T., Austin-Phillips, S., Amasino, R. M., & Krysan, P. J. (2007). The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res, 120*(1), 157-165. doi:10.1007/s10265-006-0048-x

Xie, K., Zhang, J., & Yang, Y. (2014). Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant, 7*(5), 923-926. doi:10.1093/mp/ssu009

Yin, H., Zhang, X., Liu, J., Wang, Y., He, J., Yang, T., . . . Gong, Z. (2009). Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in Arabidopsis. *Plant Cell, 21*(2), 386-402. doi:10.1105/tpc.108.061549

Zacharaki, V., Benhamed, M., Poulios, S., Latrasse, D., Papoutsoglou, P., Delarue, M., & Vlachonasios, K. E. (2012). The Arabidopsis ortholog of the YEATS domain containing protein YAF9a regulates flowering by controlling H4 acetylation levels at the FLC locus. *Plant Sci, 196*, 44-52. doi:10.1016/j.plantsci.2012.07.010

TWO ARABIDOPSIS YEATS DOMAIN PROTEINS FACILITATE DNA DAMAGE REPAIR VIA THEIR HISTONE BINDING DOMAINS.

ABSTRACT

Accurate repair of DNA double-strand breaks (DSBs) is required to maintain genomic integrity and overall health. DSB repair occurs in the context of chromatin and is mediated by the detection, deposition, and removal of a wide array of post-translational modifications. While various chromatin marks have already been implicated in DSB repair, the factors, pathways, and mechanisms through which various chromatin modifications influence DSB repair is not completely understood. In this study, we demonstrate that YAF9A and YAF9B, two homologs containing the YEATS histone acylation reader module, are required for DNA repair in a chromatin-mediated manner. We showed that, while the constitutively expressed YAF9A and DSB-induced YAF9B proteins display partially redundant roles in plant development and genotoxic stress tolerance, they have largely non-redundant roles in DSB repair via HR-like pathways, suggesting independent DNA repair roles for YAF9A and YAF9B. Using proteomic and mutagenesis experiments to gain mechanistic insights into YAF9B, we showed that YAF9B is a DSB-specific component of the NuA4 histone acetylation complex and its YEATS domain is crucial for its role in repair. Together, our functional and mechanistic characterization links YAF9 proteins to HR-like DSB repair via histone acetylation complexes and the YEATS domain.

INTRODUCTION

Organisms are constantly subjected to different kinds of DNA damage via exposure to various endogenous and exogenous DNA damaging agents(Ciccia & Elledge, 2010; Manova &

Gruszka, 2015). To repair these lesions and maintain genomic integrity, organisms have evolved cellular pathways collectively termed as the DNA damage response (DDR)(Ciccia & Elledge, 2010; Hu, Cools, & De Veylder, 2016). Of these DNA damages, double strand breaks (DSBs) are quite deleterious(Aguilera & Garcia-Muse, 2013) as they can lead to chromosomal abnormalities if repaired improperly(Rodgers & McVey, 2016). DSBs are repaired via two main pathways: the error-prone nonhomologous end joining (NHEJ) pathway and the error-free homologous recombination (HR) pathway(Scully, Panday, Elango, & Willis, 2019). DNA DSBs, and subsequent repair, occur in the context of the chromatin landscape, which can be modified by the exchange of canonical histones with variant forms, as well as by the post-translational modifications of histones (such as acetylation, methylation, phosphorylation, ubiquitylation, SUMOylation) and cytosine methylation of DNA (Jeggo, Downs, & Gasser, 2017; Scully et al., 2019). Crosstalk between multiple chromatin-modifying complexes and their targets result in a dynamic and diverse chromatin landscape and influences all DNA-associated processes. The involvement of various chromatin modifications in serving as recruitment signals after DNA damage, as well as facilitating a repair-conducive environment is emerging across organisms (Clouaire et al., 2018; Drury et al., 2012; Van & Santos, 2018). For example, the serine phosphorylation of histone variant H2A.X is a crucial modification that marks the DSB site and plays important roles in signaling the DNA damage response(Fernandez-Capetillo, Lee, Nussenzweig, & Nussenzweig, 2004; Friesner, Liu, Culligan, & Britt, 2005; Georgoulis, Vorgias, Chrousos, & Rogakou, 2017; Siddiqui, Francois, Fenech, & Leifert, 2015; Srivastava, Gochhait, de Boer, & Bamezai, 2009), which subsequently results in the transcriptional regulation (coordinated by transcriptional master regulators: p53 in mammals(Hafner, Bulyk, Jambhekar, & Lahav, 2019) and SOG1 in *Arabidopsis*(Bourbousse, Vegesna, & Law, 2018; Ogita et al., 2018;

K. O. Yoshiyama, 2016)) of genes required for key DDR processes like cell cycle, DNA repair and cell death. The important and diverse functions of γH2A.X, and other damage-associated histone modifications, highlight the significance of chromatin in the DSB repair.

Histone acetylation is a dynamic chromatin mark linked with transcription and various aspects of DNA damage repair. The YEATS (Yaf9, ENL, AF9, Taf14, Sas5) domain is a conserved histone acetylation reader module. Yeast contains three YEATS domain proteins (Yaf9, Taf14, and Sas5) and humans contain four YEATS domain proteins (GAS41, ENL, YEATS2 and AF9)(Schulze, Wang, & Kobor, 2009). Characterization of these YEATS domain proteins showed that the YEATS domain binds to a wide range of histone acylations, including acetylation, propionylation, butyrilation, crotonylation and succinylation (Andrews et al., 2016; Barnes, English, & Cowley, 2019; Klein, Ahmad, et al., 2018; Klein, Vann, et al., 2018; Y. Li et al., 2016; Y. Li, Zhao, Chen, & Li, 2017; Schulze et al., 2009; Y. Wang et al., 2018; Zhang et al., 2016; D. Zhao et al., 2016; D. Zhao, Li, Xiong, Chen, & Li, 2017), with higher affinities for bulkier acyl marks accomplished through diverse mechanisms(Andrews et al., 2016; Y. Li et al., 2016; Y. Wang et al., 2018; Zhang et al., 2016; D. Zhao et al., 2017). The YEATS domain binding channel is long and flat, therefore allowing recognition of longer and bulkier non-acetyl acylation(Y. Li et al., 2016), which have been linked to transcriptional activation to a similar or greater extent than acetylation(Goudarzi et al., 2016; Kebede et al., 2017; Sabari et al., 2015). In *Arabidopsis*, there are only two YEATS domain proteins, YAF9A and YAF9B(Bieluszewski et al., 2015; Crevillen et al., 2019; Zacharaki et al., 2012), both of which are homologs of yeast Yaf9 and human GAS41.

The yeast Yaf9 and human GAS41 are a part of the NuA4-SWR1 and TIP60-SCRAP chromatin remodeling complexes respectively. In the yeast *Saccharomyces cerevisiae*, SWR1-C and NuA4 complexes act in conjunction and influence H2A.Z and histone acetylation chromatin

states. SWR1-C is an ATP-dependent chromatin remodeling complex responsible for deposition of the histone variant H2A.Z. NuA4 is a histone acetyltransferase responsible for the acetylation of histones H4, H2A, and H2A.Z(Setiaputra et al., 2018). While SWR1-C comprises 14 subunits, organized around the central ATPase Swr1(Nguyen et al., 2013), the NuA4 comprises 13 subunits and organized around the central assembly platform containing Tra1 and Eaf1(Setiaputra et al., 2018). The two complexes share a four-subcomponent module containing Yaf9(Lu, Levesque, & Kobor, 2009). In mammals, homologs of the yeast NuA4 and SWR1 subunits form a hybrid complex called TIP60, which can acetylate histones H2A and H4, as well as exchange H2A with H2A.Z. Another complex in humans, SRCAP is a homolog of the yeast SWR1 and is responsible for H2A.Z deposition into chromatin. Like yeast, the TIP60 and SRCAP complexes share the histone acylation reader GAS41. In line with the role of histone acetylation and H2A.Z in stress responses, mutations in subunits of the yeast NuA4/SWR1c and mammalian TIP60/SRCAP complexes have shown DNA damage repair defects(Lu et al., 2009; Squatrito, Gorrini, & Amati, 2006). In plants, most of the yeast NuA4 and SWR1 subunits are conserved(Aslam, Fakher, Jakada, Cao, & Qin, 2019; Espinosa-Cores et al., 2020). Recent proteomic analyses performed with *Arabidopsis* homologs of SWR1 and NuA4 subunits have revealed that the components form yeast-like complexes(Y. X. Luo et al., 2020; Tan et al., 2018) with a few exceptions (unique plant specific subcomponents like MBD9(Y. X. Luo et al., 2020; Potok et al., 2019; Sijacic, Holder, Bajic, & Deal, 2019) and plant specific interactions like association of TRA1 with SWR1(Y. X. Luo et al., 2020; Potok et al., 2019)). Functional characterization of the *Arabidopsis* NuA4 and SWR1 subunits suggest roles in plant development, flowering, and stress response. Similar to human and yeast complexes, mutants of select subunits of AtSWR1 (*pie1:* homolog of yeast ATPase Swr1; *arp6* and *swc6:* non-catalytic subunits) show increased sensitivity to DNA damage,

meiotic abnormalities, and decreased somatic homologous recombination (SHR)(Rosa, Von Harder, Cigliano, Schlogelhofer, & Mittelsten Scheid, 2013). However, a more systematic analysis for the roles and mechanisms of both NuA4 and SWR1 complexes in DNA repair is yet to be performed in *Arabidopsis.*

Functional characterization of the *Arabidopsis* YEATS domain proteins, YAF9A and YAF9B suggest roles in plant development. Previous studies have shown that YAF9A is constitutively expressed throughout the plant and *yaf9a-1* mutants display early flowering phenotypes, while YAF9B is detected only in young flowers and roots, and *yaf9b-2* mutants behave like wild-type(Crevillen et al., 2019; Zacharaki et al., 2012). The double mutant *yaf9a-1,b-2* displays pleiotropic developmental phenotypic alterations (conspicuous early flowering, accelerated senescence and reduced organ and plant size due to lower endoreduplication levels) and displays mis-regulation of a wide variety of genes involved in cell size, growth regulation, systemic acquired response, and flowering time regulation, suggesting that the YAF9 proteins have partially redundant functions in developmental responses(Bieluszewski et al., 2015; Crevillen et al., 2019; Zacharaki et al., 2012). Concurring with the YEATS domain reader function in yeast and mammals, *in vitro* experiments show that AtYAF9A and AtYAF9B are localized to the nucleus and are histone H3 and H3 acetylation readers(Crevillen et al., 2019). While the interaction of YAF9A with subcomponents of the SWR1/NuA4 complexes has been recently confirmed(Y. X. Luo et al., 2020; Potok et al., 2019), the interactions of YAF9B remain unexplored. Moreover, despite the known involvement of key SWR1 sub-components in DNA repair(Rosa et al., 2013), the roles of the histone reader proteins, YAF9A and YAF9B in DNA damage repair and the mechanisms through which they act, remain unexplored. Given the strong connections of H2A.Z

and histone acetylation with stress response, it is important to investigate the roles of YAF9 subunits in DNA repair to understand the role of chromatin in mediating DNA repair.

In this work, we have shown that YAF9B is specifically induced in response to DNA double strand breaks and its expression is dependent on the palindromic motif CTT$(N)_7$AAG in the YAF9B promoter which is bound by SOG1—the master transcriptional regulator of the DNA damage response in plants. Consistent with the roles of SWR1 subunits in DNA damage repair(Rosa et al., 2013), *yaf9* mutants are hypersensitive to DNA damage-inducing agents in a redundant manner. Moreover, *yaf9* single mutants show strong defects in somatic homologous recombination (HR)-like repair, which are only slightly more aggravated in the double mutants, suggesting independent roles for YAF9A and YAF9B in HR-like repair. Consistent with the affiliation of YEATS domain proteins to chromatin remodeling complexes, proteomic experiments using YAF9B epitope-tagged lines after DNA damage show that YAF9B specifically interacts only with NuA4 subunits while recently published proteomic experiments show that YAF9A interacts with both NuA4 and SWR1 subunits(Y. X. Luo et al., 2020; Potok et al., 2019; Sijacic et al., 2019; Tan et al., 2018). This suggests that YAF9B is a DNA damage specific component of NuA4 complex with a role in DNA repair. Plants with alanine substitutions of key histone interacting amino acids in YAF9B are hypersensitive to DNA damage-inducing agents, emphasizing the importance of the reader function of YAF9B's YEATS domain in DNA repair. Overall, we show that both YAF9A and YAF9B are required for DNA repair and delve into the molecular mechanism mediating YAF9B's role in DNA damage repair via the YEATS reader function and chromatin remodeling complex.

RESULTS AND DISCUSSION

YAF9B induction is DSB-specific, requires the SOG1 motif, and is highest in meristematic zones.

Analysis of previously published transcriptomic data(Bourbousse et al., 2018) revealed that the expression pattern of *YAF9B* is similar to many other genes important for DNA repair(Bourbousse et al., 2018)–it is a direct target of SOG1 and shows low expression under control conditions, but is strongly induced by γ-IR in a SOG1-dependent manner (**Figure 3-1A**). In contrast, *YAF9A*, its closest homolog, is highly expressed under control conditions, and shows no induction after γ-IR or dependance on SOG1 (**Figure 3-1A**). To reveal where *YAF9A* and *YAF9B* are expressed, and to investigate the *cis*-regulatory elements and types of DNA damage controlling *YAF9B* expression, *β-glucuronidase* (*GUS*) reporter constructs driven by the endogenous *YAF9* promoters were generated. Consistent with previous RT-PCR data showing widespread expression of *YAF9A* in different plant organs in non-stressed conditions(Crevillen et al., 2019; Zacharaki et al., 2012), several independent *pYAF9A::GUS* reporter lines (ins#1-3) showed strong expression across all tissues in both 8-day old and 14-day old seedlings irrespective of γ-IR treatment (**Figure 3-1B and Figure 3-S1A)**. In contrast, the *pYAF9B::GUS* reporter lines (ins#1-2) only showed signal after γ-IR, with the strongest *GUS* expression detected in the shoot apex and root primordia of 8-day and 14-day old seedlings **(Figure 3-1B and Figure 3-S1B).** The expression patterns observed for the *pYAF9::GUS* reporter lines were validated *in vivo* via mRNA-seq experiments using dissected tissues from similarly staged wild-type seedlings either with or without exposure to γ-IR **(Figure 3-1C)**. While *YAF9A* was uniformly expressed, *YAF9B* expression was induced by γ-IR and was the highest in samples containing meristematic tissues (*i.e.* the shoot apex and root+ hypocotyl; **Figure 3-1C)**. These expression analyses demonstrate that in addition to being induced by DNA damage, *YAF9B* expression is also tissue-specific,

closely mimicking the expression pattern of SOG1(K. O. Yoshiyama et al., 2013) and other DNA repair proteins(Ogita et al., 2018; Seo, Maeda, & Hiratsuka, 2007).

Next, the role of the SOG1 *cis*-regulatory motif in regulating *YAF9B* expression was assessed using the *pYAF9B::GUS* reporter. First, to confirm that expression of this reporter is dependent on SOG1, like the endogenous *YAF9B* gene **(Figure 3-1A)**, it was crossed into two *sog1* mutant backgrounds (*sog1-1*(K. Yoshiyama, Conklin, Huefner, & Britt, 2009) and *sog1-101*(Ogita et al., 2018)). As expected, the *pYAF9B::GUS* reporter was induced in meristematic tissues after γ-IR in the wild-type co-segregants from these crosses and this expression was lost in both the *sog1* mutant backgrounds **(Figure 3-1D, E)**, recapitulating the *in vivo* expression pattern of *YAF9B*. The role of the SOG1 *cis*-regulatory motif was then assessed using mutant versions of the *pYAF9B::GUS* reporter that harbor either an inversion of conserved bases in the SOG1 motif **(Figure 3-1F)** or a deletion that alters the spacing within the SOG1 motif **(Figure 3-1G)**. For each construct, several independent lines were tested, and in all cases, mutations within the SOG1 motif resulted in strongly reduced *GUS* expression in both 8-day old and 14-day old seedlings **(Figure 3-1F, G and Figure 3-S1C, D)** as compared to the native *pYAF9B::GUS* reporter **(Figure 3-1B and Figure 3-S1B)**. These findings demonstrate that the SOG1 *cis*-regulatory motif is critical for *YAF9B* induction in meristematic tissues after DNA damage.

To determine whether *YAF9A* expression is responsive to other genotoxic stresses and whether *YAF9B* induction is specific to DSBs, or if it is also induced by other types of DNA damage, expression of the *pYAF9::GUS* reporters were assessed after exposure to a variety of different DNA damaging conditions. For *YAF9A*, the GUS staining patterns **(Figure 3-S1 E)** and endogenous expression levels remained largely unaffected under all conditions **(Figure 3-S1F)** demonstrating that *YAF9A* is not responsive to genotoxic stress. For *YAF9B*, the GUS staining was

again specific to shoot apex and root tips and was only observed in response to DSB inducing agents (γ-IR and bleomycin (Manova & Gruszka, 2015; Takahashi et al., 2019)) and not in response to other stresses (crosslinking(Bouyer et al., 2018; Gomez, Falcone Ferreyra, Sheridan, & Casati, 2019; Manova & Gruszka, 2015; Poklar et al., 1996), replication(Saban & Bujak, 2009; Takahashi et al., 2019) and environmental stress(Hong et al., 2017; Mahapatra & Roy, 2021; Takahashi et al., 2019)) **(Figure 3-1H)**. Consistent with these results, increased expression of endogenous *YAF9B* was also specific to DSB-inducing agents **(Figure 3-S1 F)**. For all the aforementioned genotoxic stresses, the proper conditions were verified by assessing the induction of previously characterized stress-responsive genes (Bohmdorfer et al., 2011; Bourbousse et al., 2018; Bouyer et al., 2018; Mintoff, Rookes, & Cahill, 2015; Ryu et al., 2019; Sakuma et al., 2006; L. Wang et al., 2020; Yu et al., 2020) **(Figure 3-S1 G)**. Taken together these studies elucidate the different expression patterns of YAF9 proteins and demonstrate DSB-specific expression patterns for YAF9B. Overall, the DSB-specific, SOG1-dependent *YAF9B* induction and localization to meristematic tissues, like that of several other DNA repair factors(Hirakawa, Hasegawa, White, & Matsunaga, 2017; Ogita et al., 2018; Takahashi et al., 2019), suggests important roles for YAF9B in DNA repair.

yaf9 mutants are sensitive to genotoxic stress

To investigate the roles, and possible redundancy, of YAF9A and YAF9B in DNA repair, several *yaf9* single and double mutants were characterized. First, the previously described *yaf9a-1*(Alonso et al., 2003; Bieluszewski et al., 2015; Crevillen et al., 2019; Zacharaki et al., 2012) and *yaf9b-2*(Alonso et al., 2003; Bieluszewski et al., 2015; Crevillen et al., 2019) single mutants were used to generate the *yaf9a-1,b-2* double mutant, reproducing the previously described

developmental defects, including reduced organ and plant size (Bieluszewski et al., 2015; Crevillen et al., 2019; Zacharaki et al., 2012) (**Figure 3-S2A**). Given the strong phenotype of the *yaf9a-1,b-2* double, a second, potentially weaker allele of *yaf9b* was obtained. This allele, *yaf9b-3*, harbors a T-DNA upstream of the *YAF9B* gene **(Figure 3-2A)** and when combined with *yaf9a-1*, the *yaf9a-1,b-3* double displays minimal developmental defects (**Figure 3-S2A**). Further characterization of the *yaf9* alleles via mRNA-seq experiments (with or without γ-IR treatments) revealed that, unlike the *yaf9a-1* and *yaf9b-2* alleles, which are result in mis-spliced and/or truncated transcripts, respectively **(Figure 3-2A and Figure 3-S2D)**, the *YAF9B* transcript is lowly expressed in the *yaf9a-1,b-3* in both mock and γ-IR conditions (**Figure 3-2A and Figure 3-S2B**). Amplification and sequencing of the YAF9B transcript in *yaf9b-3* and *yaf9a-1,b-3* confirmed the presence of a full length, wild-type transcript **(Figure 3-S2C)**, demonstrating that this allele is a knock-down rather than a loss of function mutant, explaining its weaker developmental phenotype (**Figure 3-S2A**). In addition, fine mapping of the T-DNA insertion site in *yaf9b-3* revealed that, it is located just downstream of SOG1 motif in *YAF9B* promoter (**Figure 3-2A**), suggesting that the proximity of this motif to the gene is also critical for DNA damage-dependent induction of *YAF9B*. Together, these *yaf9* alleles provide a series of mutants with variable strengths and expression patterns that can be leveraged to assess the roles of YAF9A and YAF9B in DNA repair.

One commonly used method to identify defects in DNA repair is the "true leaf" assay in which the accumulation of DNA damage causes cell cycle arrest or cell death in the apical meristem of young seedlings, resulting in the development of deformed or missing true leaves(Rosa & Scheid, 2014a, 2014b; Rosa et al., 2013). Compared to both wild-type (Col-0) and no treatment controls, a higher proportion of the *yaf9a-1,b-2* double mutant plants in this assay showed true leaf defects after exposure to 100Gy of γ-IR (**Figure 3-2B**). Notably this result was

consistent across several independent experiments (**Figure 3-2B, C**) and was dose dependent (**Figure 3-S2E**). The other single and double *yaf9* mutants, however, showed defects similar to the wild-type control (**Figure 3-2B**). The lack of defects in *yaf9a-1,b-3* suggests that the residual *YAF9B* expression in the *yaf9a-1,b-3* mutant (**Figure 3-S2C**) is sufficient to protect against DNA damage. To confirm that the damage sensitivity observed in the *yaf9a-1,b-2* mutant is due to disruption of these *YAF*9 genes, complementation assays were conducted. To avoid the fertility defects observed in the homozygous *yaf9a-1,b-2* double mutant, FLAG or GFP tagged *YAF9A* or *YAF9B* genes, driven by their endogenous promoters, were introduced into *yaf9a-1,b-2* populations segregating for either *yaf9b-2* or *yaf9a-1*, respectively, as diagramed in (**Figure 3-S2F**). After selecting for homozygous transgenic lines, and genotyping for the *yaf9a-1* and *yaf9b-2* alleles, true leaf assays were conducted using the YAF9A and YAF9B tagged lines in the *yaf9a-1,b-2* mutant background (**Figure 3-2C**). These assays demonstrated that both the GFP and FLAG tagged YAF9A and YAF9B proteins are able to largely rescue the DNA damage defects observed in the *yaf9a-1,b-2* mutant, as is evidenced by the reductions in seedlings with deformed or missing true leaves after γ-IR exposure (**Figure 3-2C**). Altogether, these findings demonstrate that YAF9A and YAF9B act redundantly to promote plant growth after genotoxic stress.

YAF9A and YAF9B are both required for homologous recombination after γ-IR

To determine whether YAF9A and YAF9B are required for HR-like repair of DSBs, two previously described reporters were utilized (Puchta & Hohn, 2012; Roth et al., 2012) (**Figure 3-3A**). Both reporter constructs contain a *GUS* gene that is nonfunctional until a DSB is introduced by an *I-SceI* restriction enzyme present on a separate inducer construct and then repaired by single strand annealing (SSA), in the case of the DGU.US reporter, or by synthesis dependent strand

annealing (SDSA), in the case of the IU.GUS reporter (**Figure 3-3A**). To assess defects in DSB repair, the reporter (DGU.US and IU.GUS) and inducer (I-SceI) constructs, each in the *yaf9a-1* or *yaf9b-2* mutant background, were crossed (see **Figure 3-S3A** for crossing scheme) and the numbers of blue sectors per seedling (indicating HR-like repair) were compared to wild-type control crosses. For both the SSA and SDSA reporters, fewer blue sectors were observed in the *yaf9a-1* and *yaf9b-2* mutants compared to wild-type controls (**Figure 3-3B**) and this phenotype was consistent across several independent crosses (**Figure 3-S3B, C**). These findings not only demonstrate roles for YAF9A and YAF9B in HR-like DNA repair, placing them alongside only one other plant chromatin readers shown to be important for this process(Lario, Ramirez-Parra, Gutierrez, Casati, & Spampinato, 2011; S. Zhao, Zhang, Yang, Zhu, & Li, 2018), but also suggest that they preform largely independent functions as defects in either gene causes defects in repair. Moreover, a reduction of blue spots in *yaf9b-2* is observed throughout the plant (**Figure 3-3B and Figure 3-S3B,C**), and not only in meristematic zones where YAF9B is highly induced. This evidence, in addition to the observed low induction of YAF9B throughout the rest of the plant (**Figure 3-1C**), suggests that YAF9B is important for DNA repair throughout the plant, with different levels of transcriptional induction.

To quantify and assess HR-like repair defects in additional mutant backgrounds, a simplified genetic approach was utilized wherein DSBs are introduced in the reporters by γ-IR, rather than the inducer construct (**Figure 3-3A**). Using this approach, repair rates at the SSA and SDSA reporters were assessed in the *yaf9a-1*, *yaf9b-2*, *yaf9b-3*, and *yaf9a-1,b-2* mutants 7 days after exposure to γ-IR and compared to wild-type controls. Across multiple independent experiments, the recombination rates (represented by number of blue sectors per seedling) were significantly reduced in the all the *yaf9* mutant plants compared to the wild-type controls in both

reporter backgrounds (Wilcox t-test **** $p \leq 0.0001$; **Figure 3-3C**). In line with *yaf9b-3* being a knock-down rather than a null allele, this mutant displayed a weaker phenotype than *yaf9b-2* (**Figure 3-3C**). By comparison, the recombination rates in the *yaf9a-1* mutant, especially for the SDSA reporter, were lower than both *yaf9b* single mutants (**Figure 3-3C**), revealing a stronger repair defect in this mutant. Finally, these repair defects were only slightly enhanced in the *yaf9a-1,b-2* double mutant (**Figure 3-3C**), confirming, in a more quantitative manner, that YAF9A and YAF9B act in a largely non-redundant manner to mediate DSB repair via HR-like mechanisms.

To confirm that the DSB repair defects observed in the *yaf9* mutants were due to reduced recombination rates, rather than, for example, epigenetic silencing of the *GUS* reporter lines(Puchta & Hohn, 2012; Ulker et al., 2012), complementation assays were conducted. For these assays, the SSA and SDSA reporter lines, in either a wild-type background or in the *yaf9* mutant backgrounds described in **Figure 3-S3D**, were crossed with the previously vetted YAF9 complementing lines (**Figure 3-2C and Figure 3-S2F**) and recombination rates were assessed in the resulting F_1 seedlings after exposure to γ-IR (**Figure 3-3**). Compared to the *yaf9a-1* and *yaf9b-2* controls, recombination rates via both SSA and SDSA were significantly increased in seedlings containing the tagged YAF9 proteins (**Figure 3-3D**), demonstrating a rescue of the recombination defects in *yaf9a-1* and *yaf9b-2* mutants. Complementation of the HR-like repair defects in the *yaf9* mutants affirms that both reporters remain competent for recombination in these mutant backgrounds and demonstrates that YAF9A and YAF9B are responsible for the observed DSB repair defects.

YAF9s participate in DSB repair in a chromatin-dependent but transcription-independent manner:

During the DNA damage response, histone acetylation readers are known to play diverse roles including regulating the expression of genes important for DNA repair and coordinating repair steps at the chromatin(Gong, Chiu, & Miller, 2016). To decipher the molecular mechanisms underlying the YAF9 involvement in DSB repair, the transcriptional defects in *yaf9* mutants were studied and the interactors of YAF9B after DNA damage were identified. To assess DNA damage-associated transcriptional defects in *yaf9* mutants, three independent mRNA-seq experiments were performed 3h after 100Gy γ-irradiation (+ γ-IR) or mock (- γ-IR) treatments (**Figure 3-S4A, Supplementary table 1**), a time when the DNA damage response is well documented (Bourbousse et al., 2018). Differentially expressed genes (|Fold change| ≥2 and p-value≤0.01) in the *yaf9* mutants compared to Col-0, were identified and visualized using volcano plots after either mock (**Figure 3-4A**) and γ-IR treatments (**Figure 3-4B, Supplementary table 2-list of DE genes**). Amongst the differentially expressed genes, to highlight the genes associated with DDR in the volcano plots, the genes are color-coded based on the path number in the DREM transcriptional model for wild-type plants that was previously generated (number of genes identified in each path are indicated below the path number) (Bourbousse et al., 2018). In the *yaf9b-2* LOF allele, no additional genes, besides YAF9B, are differentially expressed compared to the wild-type control in either the mock or γ-IR conditions, suggesting that the DSB defects in *yaf9b-2* are not due to transcriptional defects. In the other single mutants, very few genes (5 down, 11 up in mock; and 7 down, 12 up in γ-IR for *yaf9a-1* and 0 down, 3 up in mock; and 1 down, 2 up in γ-IR for *yaf9b-3*) including the respective mutated genes, are differentially expressed compared to the wild-type control. The DE genes identified in *yaf9a-1* and *yaf9b-3* are not connected to DNA damage repair.

Overall, analysis of differential expression in single mutants suggests that the independent roles of YAF9 proteins in DSB repair, especially YAF9B, are not mediated by transcriptional regulation.

Since *YAF9B* is expressed in a DSB-specific manner (**Figure 3-1**) and recent IP/MS studies have shown that the *Arabidopsis* YAF9A is a subunit of both NuA4 and SWR1 acetylation complexes (**Figure 3-5A**) (Y. X. Luo et al., 2020), it was hypothesized that YAF9B might be a part of chromatin modifying complex. To interpret the function of YAF9B by identifying its interactors, IP/MS was performed using a *pYAF9B:YAF9B:FLAG yaf9b-2* complementing line described in **Figure 3-2C**. IP/MS was performed in 8-day old *pYAF9B:YAF9B:FLAG yaf9b-2* and Col-0 (wild-type) seedlings, 6h after 100Gy γ-IR and revealed the enrichment of about 60 proteins that are more than 30-fold enriched compared to Col-0 (180 proteins more than 10-fold enriched) (**Supplementary table 5**). This includes sub-components of NuA4/SWR1 complexes like the shared components of the NuA4/ SWR1 complexes (ARP4 and SWC4), NuA4 assembly platform (EAF1A/EAF1B), NuA4 piccolo module (HAM1, EPL1B and ING2) and INO80/SWR1 shared module (RIN1, RVB2A and RVB2B). These subcomponents were also enriched in the previously published YAF9A IP/MS(Y. X. Luo et al., 2020) (**Figure 3-5B**). However, unlike in YAF9A IP/MS (Y. X. Luo et al., 2020), unique components of the SWR1 complex (PIE1, ARP6, SWC6, MBD9, SWC2) were not enriched in YAF9B IP/MS, suggesting that YAF9B might specifically associate with the NuA4 acetylation complex only. This evidence suggests the existence of a DNA damage specific, YAF9B containing NuA4 complex and DNA repair roles for YAF9 proteins via chromatin modifying complexes. Altogether, the absence of transcriptional defects in *yaf9b-2*, the DNA damage specific association of YAF9B in histone acetylation NuA4 complex and the previously described roles of histone acetylation and associated complexes in stress response(Campi, D'Andrea, Emiliani, & Casati, 2012; Fina & Casati, 2015; Fina, Masotti, Rius,

Crevacuore, & Casati, 2017; M. Luo, Cheng, Xu, Yang, & Wu, 2017; Rosa et al., 2013), link YAF9B to DSB repair in a chromatin- mediated manner.

Histone binding ability of YEATS domain is required for YAF9B function:

To investigate whether the acylation reader function of the YAF9B YEATS domain(Crevillen et al., 2019) is important in mediating its role in DNA repair, a mutagenesis approach was employed. The *Arabidopsis* YAF9A and YAF9B proteins share homology with human GAS41, in which the residues required for the YEATS domain to recognize lysine acetylation have been identified and shown to form an aromatic cage(Hsu et al., 2018; Y. Li et al., 2016) (green triangles; **Figure 3-6A**). Based on an alignment with GAS41 (**Figure 3-6A**), and a structural prediction of YAF9B (BAR tools (Winter et al., 2007)) (**Figure 3-6B**), H89, S91, F92, W111, G112 and F114 were identifed as candidate residues important for the YAF9B YEATS domain-histone interaction (**Figure 3-6A, B**). To assess the roles of these amino acids in DNA repair, alanine substitutions of the aforementioned residues, either alone or in combinations of up to four residues, were introduced into the tagged YAF9B constructs and transformed into *yaf9a-1,b-2* plants. True leaf assays were then preformed in *yaf9a-1,b-2* mutants homozygous for the either the wild-type or mutated YAF9B transgenes (**Figure 3-6C**). Consistent with previous assays (**Figure 3-2B,C**), the sensitivity in *yaf9* single mutants after γ-IR was comparable to the wild-type control, the *yaf9a-1,b-2* double mutant was more sensitive, and addition of functional YAF9B rescued the true leaf defects (**Figure 3-6C**). Amongst the single amino acid substitutions, the S91A, F92A, G112A and F114A mutants all complemented to a similar level as observed for the wild-type YAF9B control (**Figure 3-6C**), demonstrating that alone, these mutations are not sufficient to disrupt YAF9B function *in vivo*. In contrast, for all YAF9B constructs harboring the

W111A mutation, the true leaf defects remained at levels nearly equivalent to the *yaf9a-1,b-2* double mutant, with the W111A mutants in the triple and quadruple mutant combinations complementing the least (**Figure 3-6C**). The increased severity of the defects in triple and quadruple mutant combinations, compared to W111A single, suggests that these residues play a complimentary role in the YEATS domain function. Notably, the sensitivities observed in the YAF9B substitutions are in agreement with the previously identified importance of their counterparts in GAS41 (Hsu et al., 2018). Altogether, this data suggests that a functional YAF9B YEATS domain is required for its role in DNA damage. Moreover, the importance of the Tryptophan (W) in the aromatic cage for YEATS domain function in both *Arabidopsis* YAF9B and Human GAS41(Hsu et al., 2018), suggests a possible conservation of the histone-binding mechanism.

CONCLUSIONS

This study shows that two differentially expressed *Arabidopsis* YEATS domain-containing proteins, YAF9A and YAF9B, are both required for proper DSB-repair via HR-like pathways and that the role of YAF9B in DSB repair is chromatin-mediated and dependent on its YEATS domain. Transcriptomics and endogenous promoter driven GUS experiments showed that while YAF9A is ubiquitously expressed in a DNA damage-independent manner, YAF9B is conditionally induced upon DSB damage in a SOG1-dependent and tissue-specific manner (in meristematic tissues) (**Figure 3-1; S1**). Molecular DNA repair assays using *yaf9* mutant alleles of variable strengths and higher order mutants, showed that YAF9A and YAF9B have independent roles in HR-like DSB repair pathways (**Figure 3-3; S3**). Proteomic experiments showed that DSB-induced YAF9B specifically interacts with components of the NuA4 histone acetylation complex but not unique

components of the SWR1 H2A.Z depositing complex (**Figure 3-5**). Site-directed mutagenesis experiments showed that the histone binding ability of YAF9B YEATS domain is important for YAF9B's role in DNA repair (**Figure 3-6**). Altogether, this work implicates the *Arabidopsis* YAF9 proteins in DSB repair, and adds them to a very short list of chromatin readers that have been connected to DNA repair(Lario et al., 2011; S. Zhao et al., 2018).

While previous works have shown partially redundant roles for *Arabidopsis* YAF9 proteins in development (major roles for YAF9A and minor roles for YAF9B)(Crevillen et al., 2019; Zacharaki et al., 2012), our studies show that both YAF9A and YAF9B have independent roles in mediating HR-like repair. These independent roles could be attributed to various reasons. Firstly, it might be associated with their differential tissue- and DNA damage-specific expression patterns, where YAF9A is ubiquitously expressed and YAF9B is induced by DSBs in meristematic tissues (like other DNA repair proteins (K. O. Yoshiyama et al., 2013) (Ogita et al., 2018; Seo et al., 2007)). Although GUS localization experiments show induction of YAF9B in meristematic tissue (**Figure 3-1B; S1B**), transcriptomics in dissected tissues revealed that in addition to strong induction in meristematic zones, low levels of YAF9B expression are detected in other plant tissues (**Figure 3-1C**). Moreover, HR-like reporter assays in *yaf9b* mutants are defective in proper repair throughout the plant and not specifically in meristematic tissues (**Figure 3-3B; S3B,C**). This evidence suggests that the YAF9 proteins are expressed at different levels and take part in DNA repair throughout the plant. Moreover, our proteomic data showing interaction of YAF9B with NuA4-specific subcomponents after DSB damage(**Figure 3-5**), while YAF9A was previously shown to be interacting with both NuA4- and SWR1-specific subcomponents(Y. X. Luo et al., 2020; Potok et al., 2019; Sijacic et al., 2019; Tan et al., 2018), suggests DNA repair roles for YAF9B via DNA damage-specific chromatin modifying complexes. These expression patterns

and selective interactions pave way for intricate mechanisms, where YAF9A and YAF9B are a part of polymorphic and combinatorially assembled chromatin complexes, allowing them to take on highly specialized functions in a tissue-specific or a sequential manner. This is in line with previous studies in both plants(Zhou et al., 2021) and mammals(Kadoch & Crabtree, 2015; Maehara et al., 2015; Nitarska et al., 2016), where protein homologs (in chromatin complexes) regulate epigenetic landscapes and/or gene expression by taking on independent and/or tissue-specific roles. Finally, the independent roles of YAF9 proteins could be contributed by different histone binding properties of the YAF9A and YAF9B YEATS domain. Recent studies in mammalian and yeast YEATS domain proteins show that the YEATS domain binds to a broad array of histone acylations (including acetylation, propionylation, butyrylation, crotonylation and succinylation), with higher affinities to bulkier groups(Barnes et al., 2019). However in *Arabidopsis*, previous studies have tested YEATS binding ability only with select histone marks(Crevillen et al., 2019). Perhaps, further characterization of the *Arabidopsis* YEATS-histone interaction will reveal specific binding affinities and shed light on their independent roles in HR-like repair.

Altogether, this work demonstrating the role of *Arabidopsis* YEATS domain proteins in DNA repair via histone acetylation sheds light on a major unanswered question in the field of DNA damage repair- the role of chromatin in modulating DNA damage repair. This work is one of the first studies linking a histone reader to DSB repair in *Arabidopsis*, and hence a step towards understanding the role of histone modifications and their interactors in influencing DSB repair via chromatin-mediated mechanisms.

MATERIALS AND METHODS

Plant Materials:

All *Arabidopsis* mutant and transgenic lines used in this study were in the Col-0 ecotype unless specified otherwise and were either grown at 22°C under long-day conditions (16 h light, 8 h dark) unless specified otherwise. The following, previously characterized transfer DNA (T-DNA) insertion mutants were obtained from Arabidopsis Biological Resource Center (ABRC) : *yaf9a-1* (SALK_106430, (Zacharaki et al., 2012)), *yaf9b-2* (SALK_046223, (Bieluszewski et al., 2015)), and *sog1-101* (GABI_602B10, (Ogita et al., 2018)). The *sog1-1* EMS mutant (K. Yoshiyama et al., 2009), which is in a mixed Landsberg erecta (Ler)/Col background containing a *pCYCB1;1::GUS* fusion, was provided by A. Britt, Department of Plant Biology, University of California, Davis, CA. A second, uncharacterized *yaf9b* allele, *yaf9b-3* (WISCDSLOX377-380P19/CS853514, (Woody, Austin-Phillips, Amasino, & Krysan, 2007)), was obtained from ABRC. Homozygous lines were identified using PCR based genotyping using the primers listed in **Supplementary table 3**. Developmental phenotypes of plants grown in the greenhouse (22°C under long-day conditions: 16 h light, 8 h dark) shown in **Figure 3-S2A,** were captured using a digital camera. Images were processed using photopea (https://www.photopea.com/).

Generation of gateway plasmids and transgenic *Arabidopsis* plant lines:

Constructs containing *GUS* reporters or epitope tagged versions of *YAF9A* or *YAF9B* were generated using the MultiSite Gateway® Three- Fragment Vector Construction Kit (Cat # 12537-023, Invitrogen). To generate BP constructs containing either the promoter regions, genic regions (without STOP codon), or 3'UTR regions of *YAF9B* and *YAF9A*, the corresponding genomic regions were amplified using Phusion polymerase (Cat # M0530, NEB) from wild-type genomic

DNA using the primers listed in **Supplementary table 3**. The resulting PCR products were gel purified (Cat # D4001, Zymo Research) and recombined into the following pDONR entry vectors (pDONR-P4P1r for promoter regions, pDONR-221 for genic regions, and pDONRP2rP3 for 3'-UTR regions) using the Gateway BP Clonase II Enzyme kit (Cat # 11789020, Invitrogen) according to manufacturer's instructions. These sequence verified BP clones, as well as previously generated BP clones containing (pDONR221-GUS, pDONRP2rP3-3XFLAG-BLRP^{mut} and pDONRP2rP3-GFP) were then recombined with the pB7m34GW destination binary vector(Karimi, Depicker, & Hilson, 2007) using the Gateway LR clonase kit (Cat# 11791019, Invitrogen) to generate the constructs detailed in **Supplementary table 4**.

Constructs containing mutations in the SOG1 motif, in the YEATS domains of YAF9A or YAF9B were generated by performing Site directed mutagenesis (SDM) on the BP constructs pDONRP4-P1R-pYAF9B, pDONR221-YAF9A, or pDONR221-YAF9B, respectively, using the primers listed in **Supplementary table 3**. SDM was performed using a two-step method. Briefly, the original BP constructs were amplified with NEB Phusion® High-Fidelity DNA Polymerase (Cat # M0530S, NEB) and complimentary primers containing the desired mutation (listed in **Supplementary table 3**). The PCR product was digested with DpnI (Cat # R0176S, NEB) and transformed into One Shot™ MAX Efficiency™ DH5α-T1R Competent Cells (Cat # 12297016, Invitrogen). These BP plasmids were used to generate the destination vector constructs listed in **Supplementary table 4**.

After sequence verification, the destination vectors were transformed into the AGLO *Agrobacterium tumefaciens* strain, and used to create *Arabidopsis* transgenic lines in specified genetic backgrounds (**Supplementary table 4**) using floral dip method(Clough & Bent, 1998). Primary transformants for each construct were selected on half strength Linsmaier and Skoog

(LS½) medium with 25ug/ml Basta (Cat # J66186, Alfa Aesar) and genotyped for the genetic background with primers listed in **Supplementary table 3**. Subsequent generations were screened to select homozygous, single insert lines and the appropriate genetic backgrounds **(Figure 3-S2F)** that were used for the GUS staining, phenotype complementation and IP/MS experiments.

Lines containing the *pYAF9B::GUS::YAF9B 3'UTR+* construct in a *sog1* mutant background were generated by crossing the *pYAF9B::GUS::YAF9B 3'UTR* line (ins#4), which was generated in a *yaf9a-1 (het) yaf9b-2 (ho)* background, into either the *sog1-1* or *sog1-101* mutant and isolating *pYAF9B::GUS::YAF9B 3'UTR (ho) yaf9a-1(wt) yaf9b-2(wt) sog1(ho)* and a wild-type cosegregant *pYAF9B::GUS::YAF9B 3'UTR (ho) yaf9a-1(wt) yaf9b-2(wt) sog1(wt)* in F3 generation. The *pCYCB1;1::GUS* in *sog1-1* was segregated out by genotyping in the *pYAF9B::GUS::YAF9B 3'UTR(ho) yaf9a-1(wt) yaf9b-2(wt) sog1-1(ho)* and *pYAF9B::GUS::YAF9B 3'UTR (ho) yaf9a-1(wt) yaf9b-2(wt) sog1-1(wt)* lines.

GUS histochemical staining of seedlings:

Seeds were gas sterilized, resuspended in 1ml of sterilized water, stratified by incubation at 4°C for 2-3 days in the dark, sowed on LS½ media plates and grown in long day conditions (16h light, 8h dark). For gas sterilization, seeds in microcentrifuge tubes, were exposed to chlorine gas (200ml bleach and 5ml hydrochloric acid (HCl), stirred continuously) in a sealed container for 1h and then aired out for 1h to vent the chlorine gas. Whole seedlings (8-day and 14-day old for promoter driven GUS lines in **Figure 3-1** and 14-day and 19-day old for HR-like repair assays in **Figure 3-3**) were fixed in ice-cold 90% acetone by submerging them completely for 20 minutes. The seedlings were briefly rinsed by completely submerging the tissue in GUS staining buffer (50 mM NaPO4, pH 7.2, 0.2% Triton X-100) at room temperature and then incubated in GUS staining

buffer supplemented with 2 mM X-Gluc (Cat # G1281C1, GoldBio) at 37°C in the dark overnight (10-16 hours). The seedlings were cleared by sequential ethanol washes (30%, 50% and then 70% ethanol) for 1h each. The cleared seedlings were stored in 70% ethanol and imaged using stereo microscope (FisherbrandTM Research Grade Stereo Zoom Microscope) and photographed with SeBaCam14C digital camera (SEBACAM14C CMOS 5V 14mp CMOS, Laxco) using SeBaView digital imaging software (Laxco). Larger seedlings were imaged on a lightbox and photographed with a digital camera.

RNA expression experiments:

RNA isolation: For all three *yaf9* mRNA-seq experiments (see **Supplementary table 1**), seeds were chlorine gas sterilized, stratified at 4°C for 3 days in the dark and grown on LS½ media with 0.6% plant agar and 1% sucrose for 8-days in long day conditions (16h-light/ 8h-dark cycles at 22°C). The seedlings were then either mock treated or exposed to γ-IR (100Gy) at a dose rate of 8-10Gy/min using a Co60 radioactive source and then returned to long day conditions for 3 hours. Two replicates of 6-8 whole seedlings from the mock and γ-IR treated samples were collected, frozen in liquid nitrogen, and stored in -80°C freezer until total RNA extraction with Quick-RNA MiniPrep kit (Zymo Research R1055) according to the manufacturer's protocol. For the tissue dissection mRNA-seq, 6-day old seedlings, grown vertically in constant light conditions (24h-light at 22°C; dissection set) were used. Approximately 100 mock or γ-IR (100Gy) seedlings were collected, 1h30' post treatment and fixed in chilled (-20°C) 80% acetone and infiltrated under vacuum for 5' twice. The seedlings were dissected one by one under a dissection scope (VWR VistaVision Stereo Microscope) in 100% ethanol with fine forceps and collected in 2mL microcentrifuge tubes filled with 100% acetone. For the RNA extraction, the acetone was

completely aspirated prior to tissue lysis using liquid nitrogen, and the RNeasy Micro Kit (Qiagen 74004) was used according to the manufacturer's protocol. For stress treatment experiments, total RNA was extracted from 4-6 whole seedlings (treated as described later), ground to a powder in liquid nitrogen, using the Quick-RNA MiniPrep kit (Cat #R1055, Zymo Research) according to manufacturer's interactions.

mRNA-seq library preparation and sequencing: mRNAs were purified from 1-2µg of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs E7490), mRNA libraries were prepared with the NEBNext UltraII RNA Library Prep Kit (New England Biolabs E7770) and sequenced on Illumina HiSeq 2500 (50-bp single-end mode).

mRNA-seq data processing: The time course mRNA-seq data for the wild-type and *sog1-1* after mock or γ-IR (100Gy) treatments and the SOG1-FLAG ChIP-seq data were previously published(Bourbousse et al., 2018) and available in Gene Expression Omnibus (GEO) database (accession no. GSE112773). For all mRNA-seq experiments, Illumina reads were mapped to the TAIR10 genome using STAR(Dobin et al., 2013), with the following parameters: maximum number of mismatches per read = 2, minimum intron length = 20 bp, maximum intron length = 6000 bp, minimum total length of exons = 5% of read length. A summary of the read mapping is presented in **Supplementary table 1**. Downstream analyses were conducted using the HOMER suite(Heinz et al., 2010). The mapped SAM files were converted to BAM using SAMtools(H. Li et al., 2009). Tag directories were created using the makeTagDirectory from HOMER(Heinz et al., 2010). To visualize the gene expression data, UCSC browser tracks and .tdf files were generated for all mRNA-seq samples using the makeUCSCfile script (-fragLength given -norm 10000000 -style rnaseq -strand both) from HOMER(Heinz et al., 2010) and igvtools(Thorvaldsdottir, Robinson, & Mesirov, 2013) respectively. Expression values

(**Supplementary table 2A**) for all genes across samples were retrieved using the analyzeRepeat.pl script from HOMER(Heinz et al., 2010) with the -noadj and -condenseGenes options. PCA plots were generated (with expression values listed in **Supplementary table 2A**) in R-Studio using plotPCA function of DESeq2(Love, Huber, & Anders, 2014) package using experimental design parameters: design = ~batch + Condition + Group + Condition:Group. The *yaf9* mRNAseq data was corrected for batch effects using removeBatchEffect function from limma package(Ritchie et al., 2015).Differentially expressed genes (FC ≥ 2 with an FDR ≤ 0.01) based on the experimental conditions (i.e., + γ-IR vs. - γ-IR or mutant vs. wild-type) were identified using the getDiffExpression_v2_DESeq2bugFixed.pl script from HOMER with the -DESeq2(Love et al., 2014) and -repeats options. The list of differentially expressed genes in each comparison are listed in **Supplementary table 2B** and the log2FC and adjusted p-value for all genes are listed in **Supplementary table 2C**. Volcano plots were generated with the log2FC and adjusted p-values using the EnhancedVolcano package(Kevin Blighe, 2021), as shown in **Figure 3-4**. Cut-offs of log2FC >|1| and p-value < 10e-2 were used and genes were colored based on the previously identified DDR gene cluster(Bourbousse et al., 2018) they belonged to.

Gel based expression analysis of YAF9B full transcript: 1.5 µg of isolated total RNA (same as RNA used for set4 of YAF9 mRNA-seq) was utilized for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit (Cat#4374967, Life Technologies) according to manufacturer's instructions. The generated cDNA was diluted five-fold and used as a template and RT-PCR was performed with GoTaq® Green Master Mix (M7123) and primers listed in **Supplementary table 3** in a thermocycler (95°C for 3min, followed by 32 cycles of denaturation at 95°C for 30s, annealing at 57°C for 30s and elongation at 72°C for 1min, followed by 72°C for

5min and 12°C forever). The resultant PCR product was migrated on a 1.5% agarose Tris-Acetate-EDTA (TAE) gel alongside a DNA ladder and imaged on a Biorad Gel Doc.

qPCR based expression analysis after stress treatments: RT-qPCR was performed on cDNA (synthesized as described above) with Bio-Rad CFX384 Real-Time System using Luna Universal qPCR Master Mix (Cat# M3003E, NEB) according to manufacturer's instructions. Transcript levels were determined using the standard curve method and relative expression was calculated by normalizing to a control gene, *AT5G13440*. Each experiment was repeated with 2 technical replicates. The gene-specific primers used in this analysis were listed in **Supplementary table 3**.

Stress treatments:

For all treatments, seeds (Col-0 and YAF9 promoter driven GUS lines) were chlorine gas sterilized, stratified, sown onto a nylon mesh on LS½ media plates with 8g/L plant agar and grown vertically for 7 days in in long day conditions (16h-light/ 8h-dark cycles). After 7 days, the plates were split into mock and treatment sets. For the treatment set, the seedlings on nylon mesh were treated as follows: For the γ-IR and UV-C light stress, the seedlings were moved to fresh LS½ media plates with 8g/L plant agar and then exposed to 100Gy γ-IR (dose rate of 8-10Gy/minute using a Co60 radioactive source) or 6000J/m2 UV-C light (dose rate of approximately 5000J/m2/minute using a Strategene UV stratalinker). Treated seedlings were moved back to vertical long day growth chamber till sample collection. For the heat and cold treatments, the seedlings were moved to fresh LS½ media plates with 8g/L plant agar and the plates were placed vertically either in 37°C chamber (dark) or 4°C cold room till sample collection. For bleomycin, cisplatin, hydroxyurea and salt treatments, the seedlings were moved to freshly made LS½ media

plates with 8g/l plant agar supplemented with 20µg/ml bleomycin (SIGMA 15361, stock= 2mg/ml in DMSO), 50µM cisplatin (SIGMA P4394, stock = 1.67mM in water), 80mM hydroxyurea (SIGMA H8627, stock = 30mg/ml in water) or 400mM NaCl (stock = 5M NaCl in water) respectively. These plates were moved back to long day growth chamber till sample collection. For the mock treatment set, the seedlings were moved to LS½ media plates with 8g/l agar as in all cases and grown in the long day chamber, with all tissue collections in parallel to the corresponding treatment set for each stress.

For relative gene expression analysis, RNA was isolated 3h post treatment from 4-6 Col-0 whole seedlings and RT-qPCR was performed as described above in three independent experiments. For GUS localization studies, 5-15 plants of each line were collected 24h post treatment and GUS stained (See GUS histochemical staining of seedlings), in two independent experiments.

True leaf assay:

Arabidopsis seeds were chlorine gas sterilized and stratified in sterile water at 4°C in dark for 3 days. The stratified seeds were γ-IR treated at a dose rate of 8-10Gy/min using a Co60 radioactive source, sown onto LS½ plates with 1% sucrose and 6g/L agar, and then transferred to a growth chamber with 16h-light/ 8h-dark cycles at 23C. After 11-12 days, the plants were scored if they have proper true leaves, deformed true leaves or absent (one/no) true leaves. The percentage of seedlings in each category was visualized using a stacked bar chart generated using ggplot2 package(Wickham, 2016) in R-Studio. The example true leaf phenotypes for seedlings with proper true leaves, deformed true leaves and one/no true leaves shown in **Figure 3-2B** were imaged using stereo microscope (FisherbrandTM Research Grade Stereo Zoom Microscope) and photographed

with SeBaCam14C digital camera (SEBACAM14C CMOS 5V 14mp CMOS, Laxco) using SeBaView digital imaging software (Laxco).

Analysis of recombination via SSA and SDSA:

The genetic materials used for assessing recombination at the SSA and SDSA reporters were generated as diagramed in **Figure 3-S3A** and **Figure 3-S3D** and recombination rates were assessed by GUS staining (See GUS histochemical staining of seedlings). For assays utilizing the *I-SceI* inducer construct to generate DSBs, GUS staining was performed on F1 seedlings grown for 15-days in long day (16h light, 8h dark) conditions. GUS-stained seedlings were imaged on a lightbox using a digital camera and processed using photopea (https://www.photopea.com/). For assays using gamma IR to generate DSBs, 7 or 8-day old, long day (16h light, 8h dark) grown seedlings were exposed to 100Gy γ-IR and let recover for 7 or 11 days. GUS staining was performed on 14 or 19-day old seedlings. The number of blue sectors per seedlings were counted using a stereo microscope (FisherbrandTM Research Grade Stereo Zoom Microscope) using 1.5X magnification. P-values were calculated from Wilcox t-tests between the indicated genotypes on R-Studio (compare_means function in package ggpubr(Kassambara, 2020)).

IP/MS:

Arabidopsis seeds (*pYAF9B:YAF9B:3XFLAG yaf9b-2* T4 and negative control Col-0) were sterilized first in a 70% ethanol solution containing 0.05% w/v SDS for 10 minutes and then in a 90% ethanol solution for 5 minutes. The sterilized seeds were stratified at 4°C for 2-3 days in the dark and grown vertically on LS½ media plates with 8g/L agar in a growth chamber with 16h-light/ 8h-dark cycles at 23C for 8 days. Approximately 10g of 8-day old seedling tissue was frozen

per genotype, 6h after 100Gy γ-IR (treated at a dose rate of 8-10Gy/min using a Co60 radioactive source). Frozen tissue was ground in liquid nitrogen and resuspended in 50mL of lysis buffer (LB: 50mM Tris pH7.6, 150mM NaCl, 5mM MgCl2, 10% glycerol, 0.1% NP-40, 0.5mM DTT, 1µg/µL pepstatin, 1mM PMSF and 1 protease inhibitor cocktail tablet (Roche, 14696200)). The tissue was then homogenized by douncing and centrifuged at 4°C in an ultracentrifuge for 25 minutes at 13,500 rpm. Each supernatant was incubated at 4°C for 1 hour with 250µL of anti-FLAG magnetic beads (Cat # M8823, SIGMA) by gentle rotation on a HulaMixer. The FLAG beads were then washed twice with 20mL of LB and five times with 1mL of LB. For each wash, the beads were rotated at 4°C for 5 minutes. Proteins were then released from the FLAG beads during five room temperature incubations with 400µL of 3XFLAG peptide (Cat # F4799, SIGMA, 50X stock =5mg/ml in PBS) at a concentration of 100µg/mL. The eluted proteins were precipitated using trichloro acetic acid (TCA). Mass spectrometric analyses were conducted as described in (Law et al., 2010).

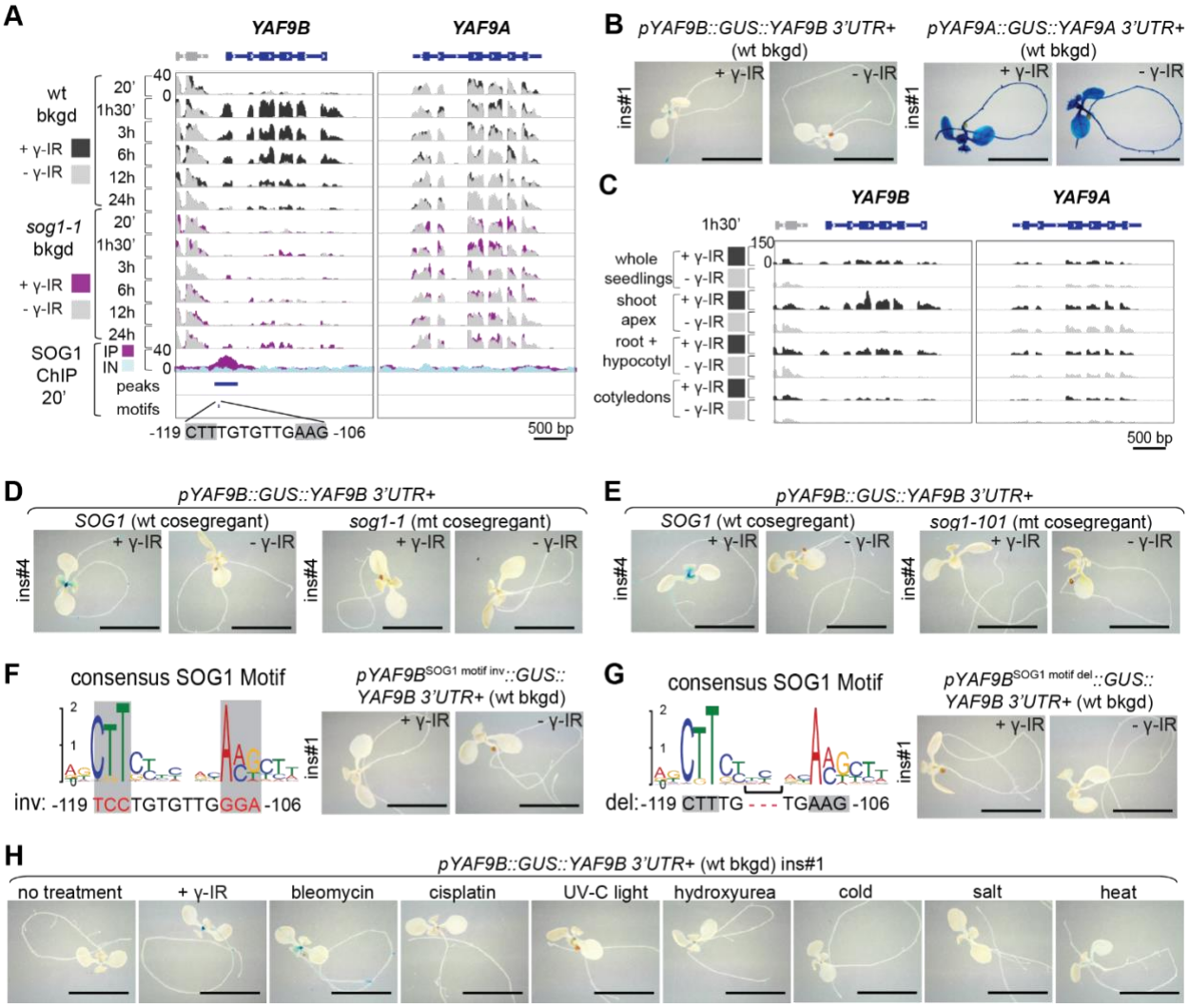


Figure 3-1. *YAF9B*, but not *YAF9A* is expressed after DNA damage in a SOG1 dependent manner.

(A) Screenshots showing the expression levels of the *YAF9B* and *YAF9A* genes in wild-type (wt) and *sog1-1* backgrounds (bkgd) at different time points, specified in minutes (') and/or hours (h), after gamma-irradiation (+ γ-IR; 100Gy) or mock (- γ-IR) treatments. SOG1 enrichment levels (+ γ-IR; 100Gy; 20') from ChIP (IP) and input (IN) samples are shown below, including the SOG1 peak region, as well as the SOG1 motif position and sequence. The *YAF9B* and *YAF9A* gene models are blue and neighboring genes are grey. The + γ-IR and - γ-IR RNA-seq tracks and IP and IN SOG1 ChIP-seq tracks are superimposed and shown on a scale of 0-40. All data is from Bourbousse et al. (2018). **(B, D-G)** Images showing the histochemical detection of β-glucuronidase (GUS) in 8-day old seedlings harboring the indicated promoter driven GUS reporter transgenes in the indicated genetic backgrounds 24h after mock (- γ-IR) or gamma-irradiation (+ γ-IR; 100Gy) treatments. Scale bars represents 5mm. **(C)** Screenshots showing the expression levels (0-150) of the YAF9B and YAF9A genes in dissected wild-type seedlings 1h30' after gamma-irradiation (+ γ-IR; 100Gy) or mock (- γ-IR) treatments. **(H)** Images showing the histochemical detection of GUS in 8-day old seedlings [pYAF9B::GUS (wt bkgd)] 24h after no treatment or exposure to gamma-irradiation (+ γ-IR; 100Gy), bleomycin (20 µg/ml), cisplatin (50µM), ultraviolet light (UV-C light,6000J/m2), hydroxyurea (80mM), salt (NaCl, 400mM), cold (4ºC) and heat (37ºC) treatments. Scale bars represent 5mm.

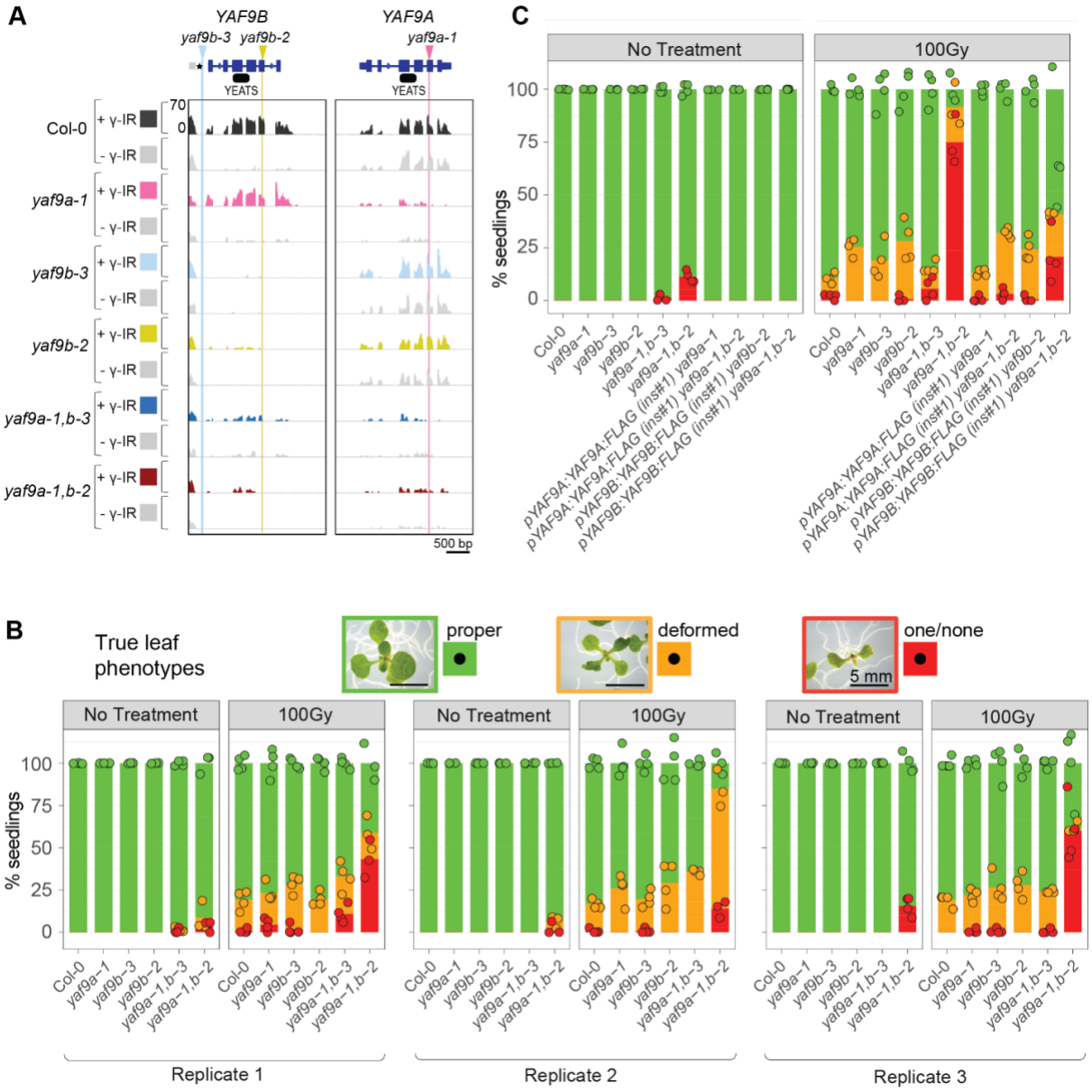
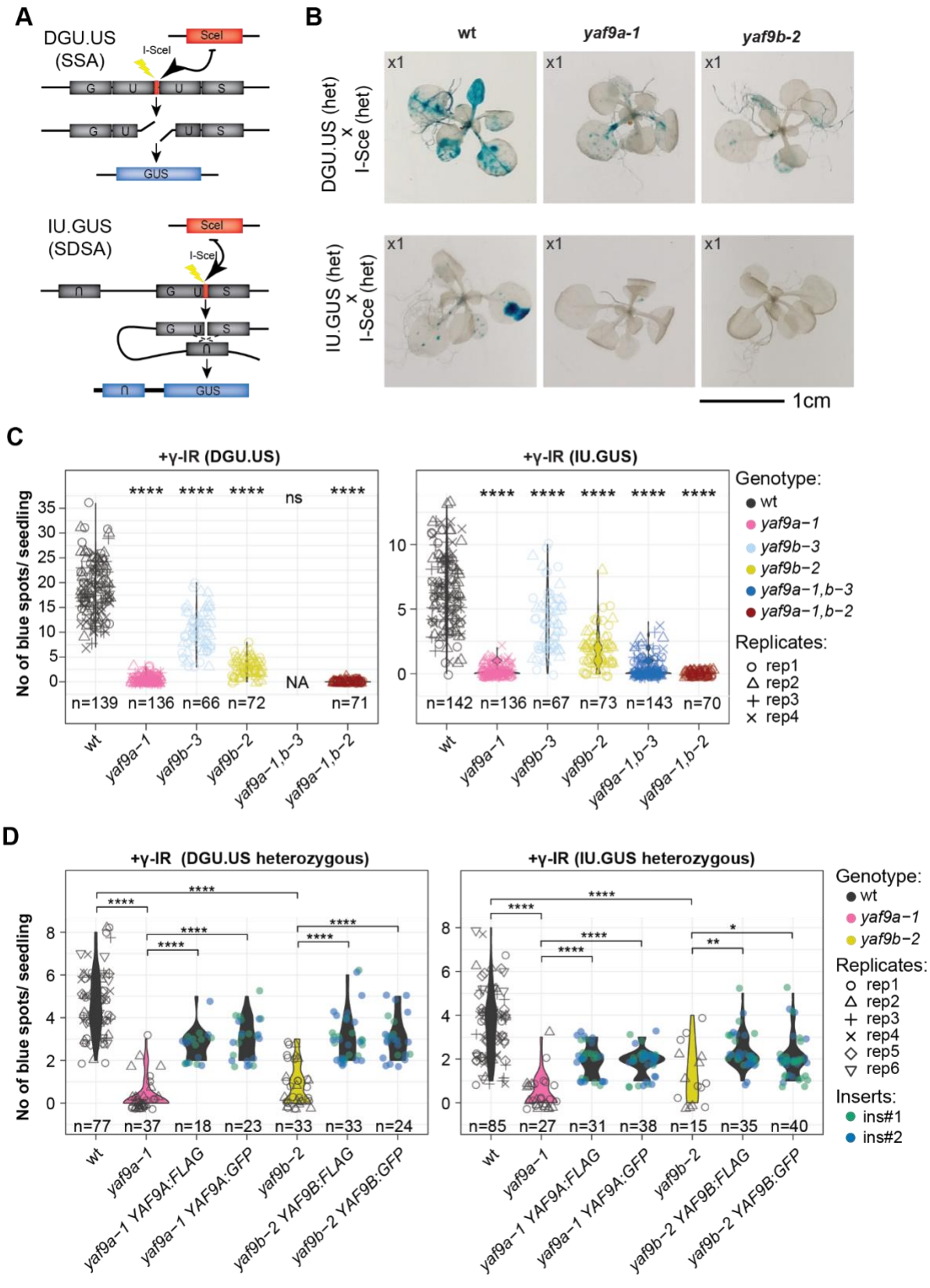


Figure 3-2. Genotoxic sensitivity in *yaf9* mutants.

(A) Screenshots showing the expression levels of *YAF9B* and *YAF9A* in wild-type (Col-0), *yaf9a-1*, *yaf9b-3*, *yaf9b-2*, *yaf9a-1,b-3* and *yaf9a-1,b-2* at 3 hours after 100Gy γ-irradiation (+ γ-IR) and mock (- γ-IR) treatments. The + γ-IR and - γ-IR tracks for each genotype are shown on a scale of 0-70. The gene models of YAF9B and YAF9A are marked with the T-DNA positions of *yaf9b-3* (blue), *yaf9b-2* (mustard) and *yaf9a-1* (pink) and the location of the YEATS domain. The location of the SOG1 motif in the *YAF9B* promoter is indicated with an Asterix (*). True leaf assay showing the percentage of seedlings with proper true leaves (green), deformed true leaves (mustard) and one/no true leaves (red) in 11-day old seedlings after treatment with 100Gy in **(B)** yaf9 single/ double mutants in 3 replicate experiments and **(C)** complementing lines alongside *yaf9* mutants. Example images of 11-day old seedlings in true leaf assay with proper true leaves (green), deformed true leaves (mustard) and no true leaves (red) are shown in **(B)**. Each circle represents a technical replicate (seeds from same seed packet grown on different plates and randomized position in the growth chamber).

Figure 3-3. YAF9A and YAF9B are required for homologous recombination after DSB damage.

(A) Transgenic reporter constructs (DGU.US and IU.GUS) in which the *β-glucuronidase (GUS)* gene was interrupted by a unique sequence including the 18-bp I-SceI recognition site. In the DGU.US construct, two direct repeats (U) of 614 bp are interrupted by a 39-bp sequence including a recognition site for the I-SceI endonuclease. Therefore, the I-SceI-induced double stranded break (DSB) can be repaired by single strand annealing (SSA) deleting 653 bp. In the IU.GUS construct, the cis homologous sequences (U) in inverse (IU) orientation may be used as templates for repair via conversion of the interrupted U containing the DSB. DSB repair by gene conversion in IU.GUS resulted in restoration of correct U and deletion of 30 bp. Targeted DSBs in the GUS reporter can be induced by I-SceI and random DSBs can be induced by exposure to genotoxic agents. In cells, where homologous recombination events occur, the restored GUS gene activity can be detected as blue spots after histochemical staining. **(B)** Representative images showing the somatic recombination events via SSA and SDSA in *yaf9* single mutants in reporter (DGU.US or IU,GUS) inducer (I-Sce) crosses. SSA/SDSA events / seedling refers to the number of blue sectors, corresponding to recombination events, in the analyzed seedling population. The cross number is indicated on the top-left corner of the image. The scale bar represents 1 cm. **(C)** Violin plots showing the somatic recombination events via SSA and SDSA in *yaf9* single and double mutants after 100Gy γ-IR. Data points for biological replicate (sibling plants) for each genotype are represented by different point shapes. The Asterix shown above represent the p-value in Wilcoxon test when compared to the wild-type (ns: p > 0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001). **(D)** Violin plots showing the complementation of SSA and SDSA defects observed in *yaf9* single mutants by YAF9 epitope tagged lines in 14-day old plants for DGU.US reporter and 19-day old plants for reporter IU.GUS. Data points for biological replicate for each genotype (independent crosses) are represented by different point shapes. Data points for rescue crosses from independent complementing epitope tagged lines (independent insertions) are shown using different colored circles. The Asterix shown above represent the p-value in Wilcoxon test (ns: p > 0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001).


A
DGU.US
(SSA)
I-SceI
SceI
GUS
IU.GUS
(SDSA)
B
wt
yaf9a-1
yaf9b-2
DGU.US (het) x I-Sce (het)
IU.GUS (het) x I-Sce (het)
x1
1cm
C
+γ-IR (DGU.US)
+γ-IR (IU.GUS)
No of blue spots/ seedling
Genotype:
wt
yaf9a−1
yaf9b−3
yaf9b−2
yaf9a−1,b−3
yaf9a−1,b−2
Replicates:
rep1
rep2
rep3
rep4
n=139 n=136 n=66 n=72 NA n=71
n=142 n=136 n=67 n=73 n=143 n=70
D
+γ-IR (DGU.US heterozygous)
+γ-IR (IU.GUS heterozygous)
rep5
rep6
Inserts:
ins#1
ins#2
n=77 n=37 n=18 n=23 n=33 n=33 n=24
n=85 n=27 n=31 n=38 n=15 n=35 n=40
yaf9a−1 YAF9A:FLAG
yaf9a−1 YAF9A:GFP
yaf9b−2 YAF9B:FLAG
yaf9b−2 YAF9B:GFP

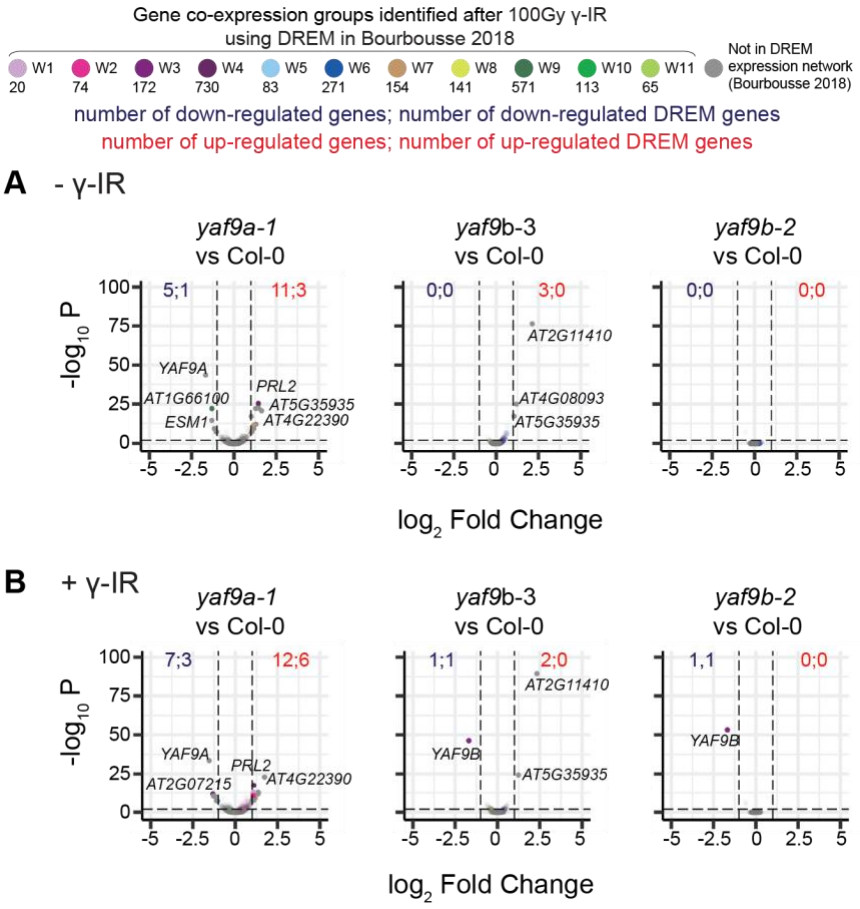


Figure 3-4. Transcriptional profiles of *yaf9* mutants with and without DNA damage.

Volcano plots showing log2 Fold change and -log10 P-value for mRNA-seq in *yaf9* mutants, 3 hours after (**A**) mock (- γ-IR) and (**B**) 100Gy γ-irradiation (+ γ-IR), identified via DESeq2. Genes are color coded according to the DREM wt path number from Bourbousse et al. (2018) (number of genes identified in each path in Bourbousse et al. (2018) are shown at the top below the path number). The opaque dots represent genes that have Fold change ≥2 and p-value≤0.01. The number of up- and down-regulated genes (and up/down-regulated genes belonging to the wild-type DREM network) clearing this cut-off for each comparison are shown in red and blue respectively.

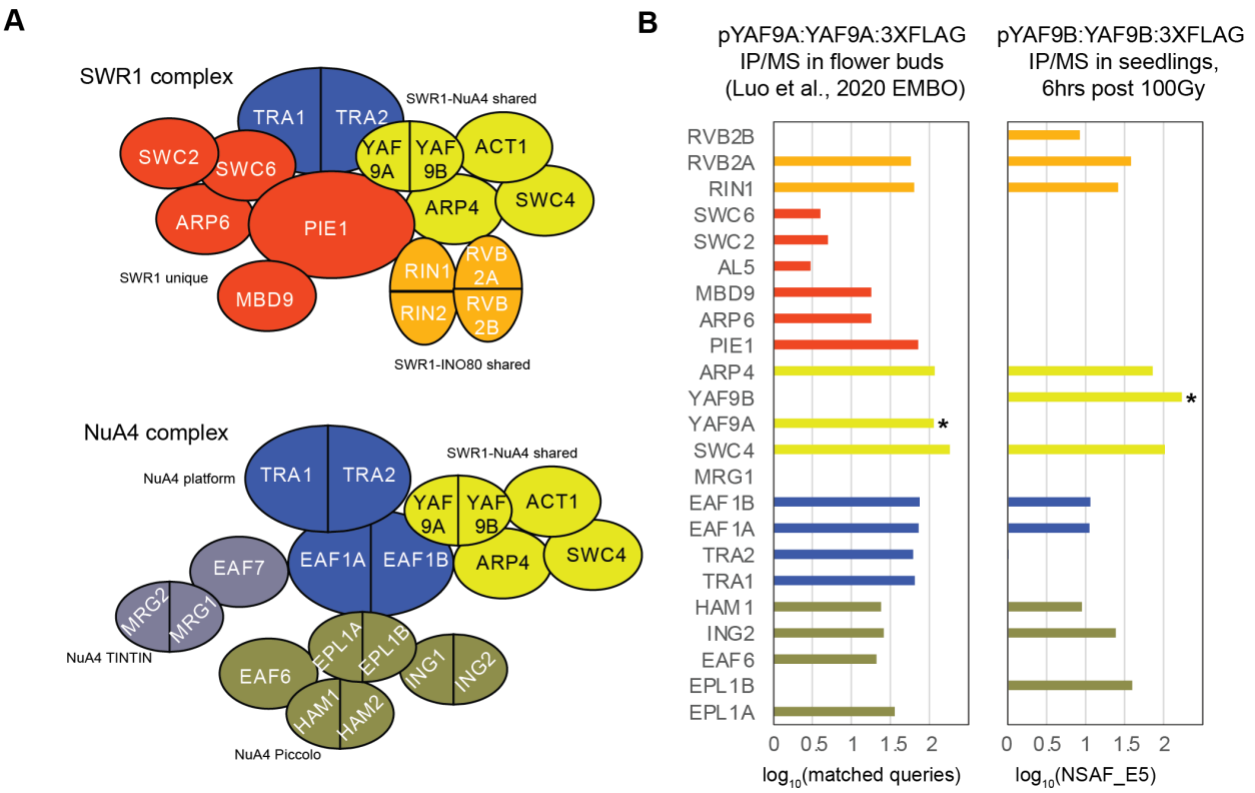


Figure 3-5. YAF9 proteins interact with subunits of histone acetylation complexes.

(**A**) Cartoons of the current understanding of the *Arabidopsis* SWR1(Y. X. Luo et al., 2020; Potok et al., 2019; Sijacic et al., 2019) and NuA4(Tan et al., 2018) complexes, color coded for sub-complexes and labeled with sub-components are shown on the left. (**B**) Log normalized protein abundance values for YAF9A IP/MS in flower buds from Y. X. Luo et al. (2020) and YAF9B IP/MS in seedlings, 6h after 100Gy γ-IR. The bait protein is highlighted with an Asterix (*) next to the data bar. Data bars are color coded according to the sub-complex the factor belongs to, as indicated in the cartoon in panel (**A**).

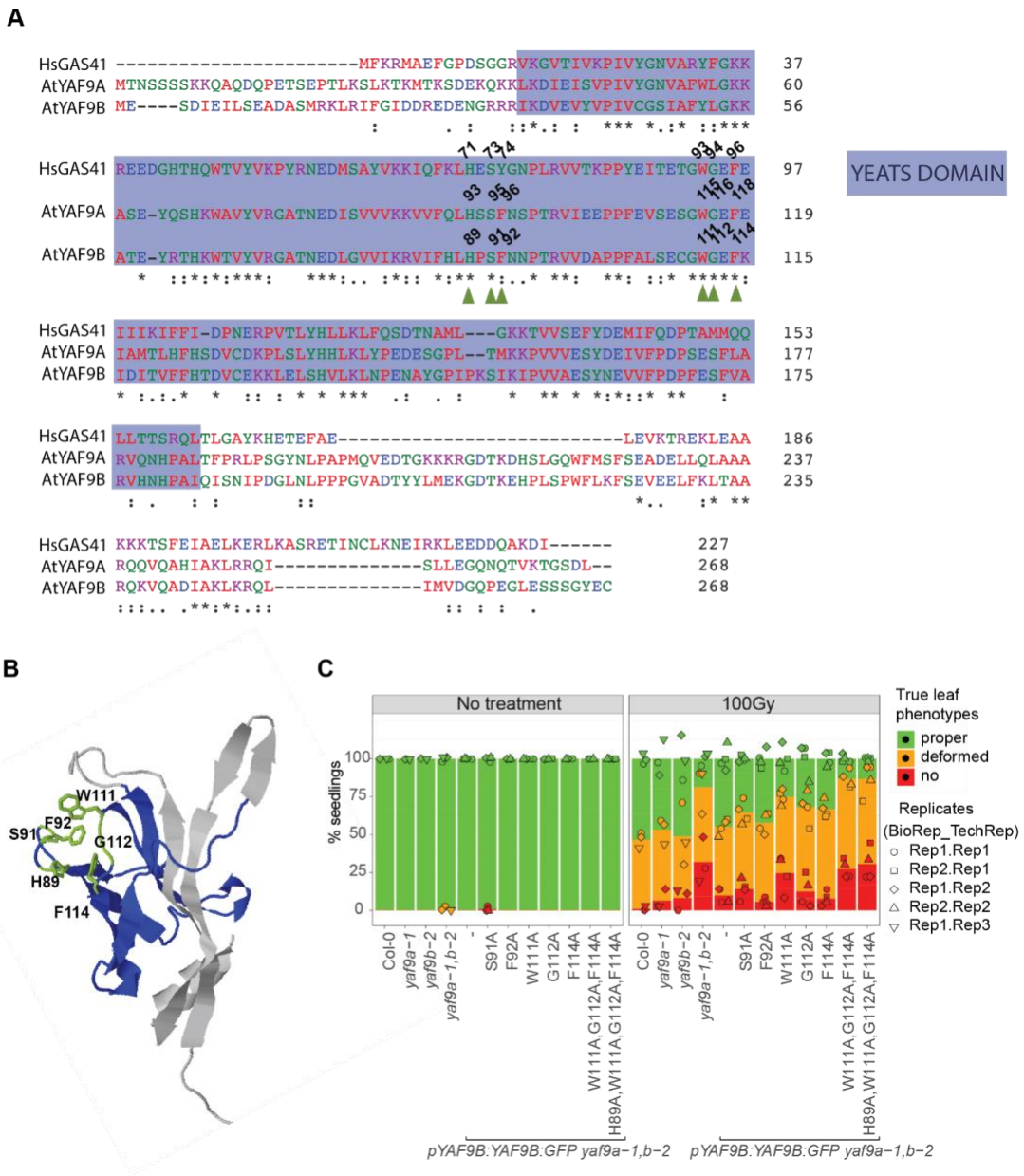


Figure 3-6. The YEATS reader domain is required for YAF9B function.

(**A**) CLUSTALW alignment(Thompson, Higgins, & Gibson, 1994) of HsGAS41, AtYAF9A and AtYAF9B. YEATS domain is highlighted in blue and residues required for the YEATS domain to recognize lysine acetylation are indicated with green arrows. (**B**) Predicted 3D-structue of YAF9B as visualized on molecular viewer on *http://bar.utoronto.ca/eplant*. The YEATS domain is colored blue and the residues predicted to interact with histone acyl marks are highlighted in green and labeled. The rest of the protein is shaded grey. (**C**) True leaf assay showing the rescue capacity of YAF9B with key acyl mark interactors mutated. The plot shows the percentage of seedlings with proper true leaves (green), deformed true leaves (mustard) and no true leaves (red) in 12-day old seedlings after treatment with 100Gy in *yaf9* single, double mutants and rescue lines. Each rescue line is represented by 1-2 biological replicates (independent insertions) and 2-3 technical replicates.

Figure 3-S1. *YAF9B*, but not *YAF9A* is expressed after DNA damage in a SOG1 dependent manner.

(A-D) Images showing the histochemical detection of *β-glucuronidase (GUS)* in 8-day and 14-day old seedlings harboring the indicated promoter driven GUS reporter transgenes in the indicated genetic backgrounds 24h after mock (- γ-IR) or gamma-irradiation (+ γ-IR; 100Gy) treatments. Scale bars represents 5mm. **(E)** Images showing the histochemical detection of GUS in 8-day old seedlings [*pYAF9A::GUS* (wt bkgd)] 24h after no treatment or exposure to gamma-irradiation (+ γ-IR; 100Gy), bleomycin (20 µg/ml), cisplatin (50µM), ultraviolet light (UV-C light,6000J/m2), hydroxyurea (80mM), salt (NaCl, 400mM), cold (4ºC) and heat (37ºC) treatments. Scale bars represent 5mm. **(F)** RT-qPCR showing the relative expression levels of *YAF9B* and *YAF9A* (normalized to a control gene, *AT5G13440*), in 7-day old Col-0 (wt) seedlings 3h after exposure to the stress treatments mentioned in **(E).** The error bars represent the standard deviation between 2 technical replicates. **(G)** RT-qPCR showing the relative expression levels of *GMI1* known to be induced by γ-IR(Bourbousse et al., 2018) and bleomycin(Bohmdorfer et al., 2011); *PR1* known to be induced by UV-C(Mintoff et al., 2015); *RAD51* known to be induced by cisplatin(Bouyer et al., 2018) and hydroxyurea(Ryu et al., 2019); *DREB2A* known to be induced by salt stress(Sakuma et al., 2006); *COR15A* known to be induced by cold stress(Yu et al., 2020); and *HSFA6A* known to be induced by heat stress(L. Wang et al., 2020), in 7-day old Col-0 (wt) seedlings 3h after exposure to the stress treatments mentioned in **(E,F).**

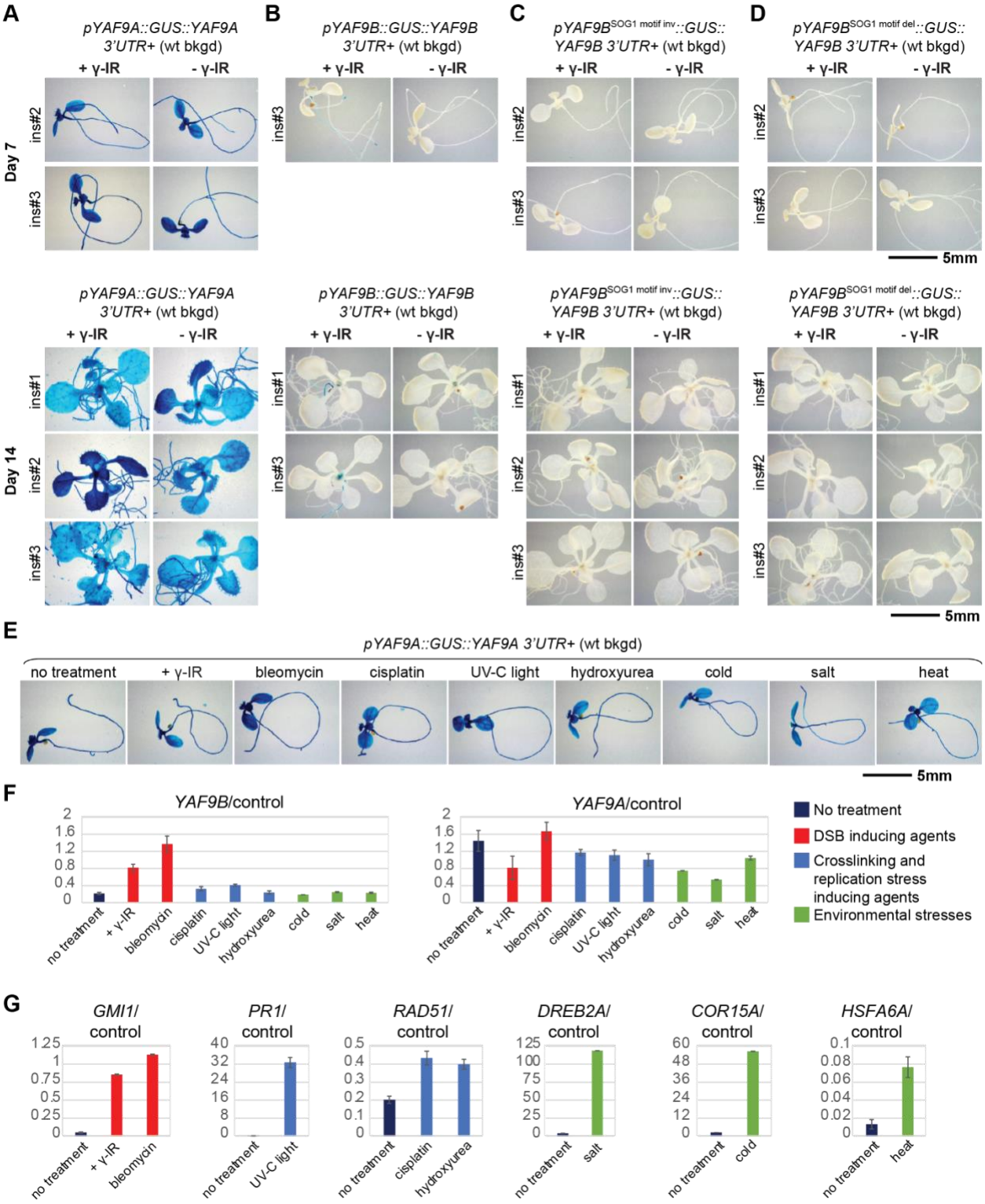

A
pYAF9A::GUS::YAF9A 3'UTR+ (wt bkgd)
+ γ-IR
- γ-IR
Day 7
ins#2
ins#3
B
pYAF9B::GUS::YAF9B 3'UTR+ (wt bkgd)
C
pYAF9B SOG1 motif inv ::GUS:: YAF9B 3'UTR+ (wt bkgd)
D
pYAF9B SOG1 motif del ::GUS:: YAF9B 3'UTR+ (wt bkgd)
5mm
Day 14
ins#1
E
pYAF9A::GUS::YAF9A 3'UTR+ (wt bkgd)
no treatment
+ γ-IR
bleomycin
cisplatin
UV-C light
hydroxyurea
cold
salt
heat
F
YAF9B/control
YAF9A/control
No treatment
DSB inducing agents
Crosslinking and replication stress inducing agents
Environmental stresses
G
GMI1/ control
PR1/ control
RAD51/ control
DREB2A/ control
COR15A/ control
HSFA6A/ control

Figure 3-S2. Characterization of *yaf9* mutants and complementing lines.

(**A**) Phenotypes of single, double *yaf9* mutants and (**F**) FLAG/ GFP tagged complementing lines in 2.5-week-old plants grown in the green house under long day growth conditions. Schematics outlining the strategy to generate epitope tagged lines complementing the single (*yaf9a-1* or *yaf9b-2*) and the *yaf9a-1,b-2* double mutant are shown in (**F**). (**B**) Screenshots showing the expression levels of *YAF9B* in wild-type (Col-0) and *yaf9a-1,b-3* at 3 hours after 100Gy γ-irradiation (+ γ-IR) and mock (- γ-IR) treatments in 4 replicates (2 biological x 2 technical replicate). The tracks for mock and γ-irradiated samples for each genotype are shown on a scale of 0-50. The gene models of *YAF9B* marked with the T-DNA positions of *yaf9b-3* (blue) and *yaf9b-2* (mustard); and the location of the YEATS domain. (**C**) RT-PCR analysis of the full length *YAF9B* transcript (807bp) expression in the shown *yaf9* mutants, 3 hours after 100Gy γ-irradiation (+ γ-IR) and mock (- γ-IR) treatments. *ACTIN2 (Act2*) was used as a control. (**D**) Screenshots showing the expression levels, along with reads mapped to *YAF9B* and *YAF9A* in wild-type (Col-0), *yaf9a-1,b-3* and *yaf9a-1,b-2* at 3 hours after 100Gy γ-irradiation (+ γ-IR) treatments. The tracks for γ-irradiated samples for each genotype are shown on a scale of 0-30. The gene models of *YAF9B* and *YAF9A* are marked with the T-DNA positions of *yaf9b-3* (blue), *yaf9b-2* (mustard) and *yaf9a-1* (pink), and the location of the YEATS domain. (**E**) True leaf assay showing the percentage of seedlings with proper true leaves (green), deformed true leaves (mustard) and one/no true leaves (red) in 11-day old seedlings after no treatment or treatment with 75Gy, 100Gy, 125Gy in *yaf9* single and double mutants. Each circle represents technical replicates (seeds from same seed packet grown on different plates and randomized position in the growth chamber).

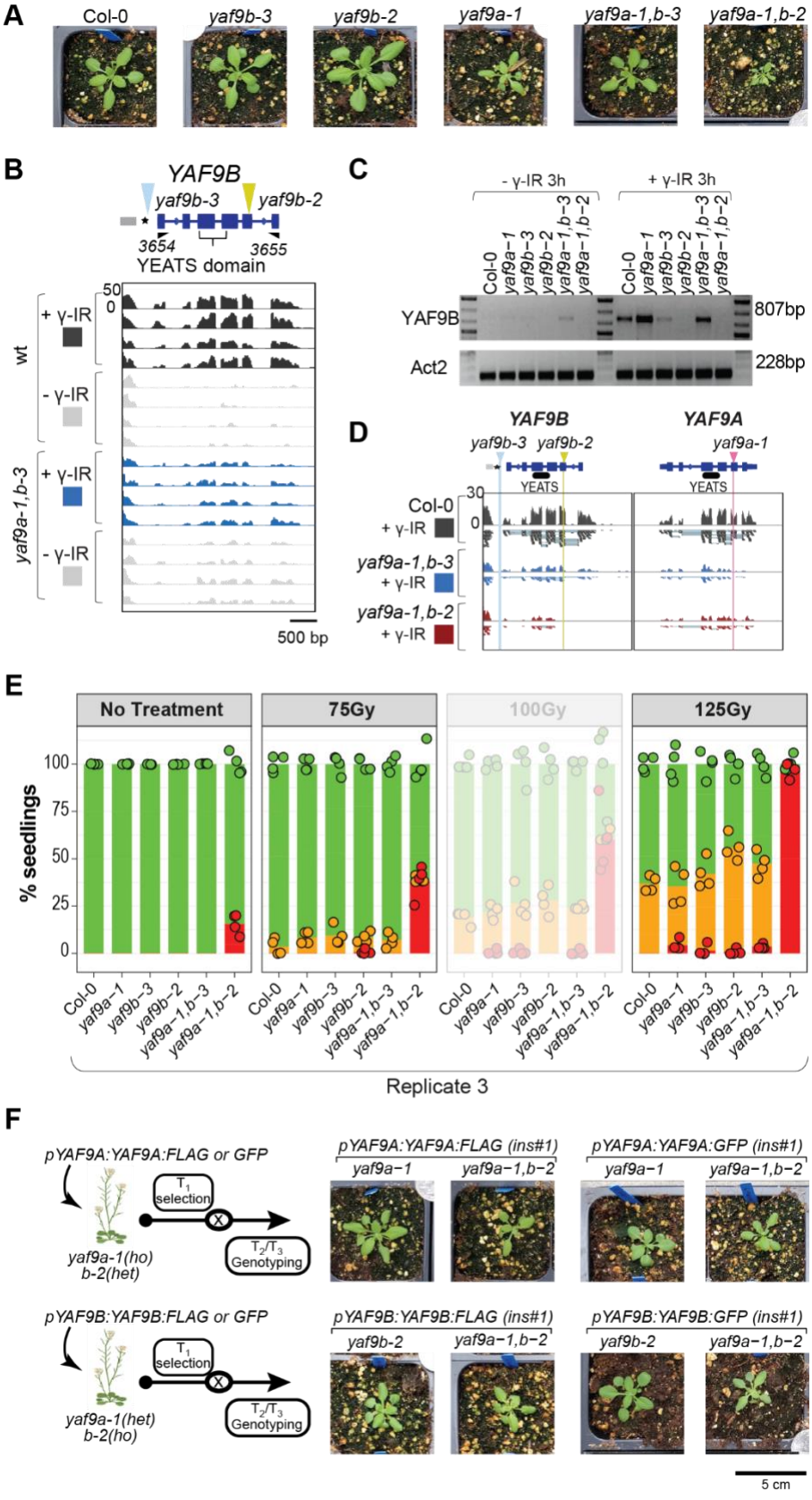
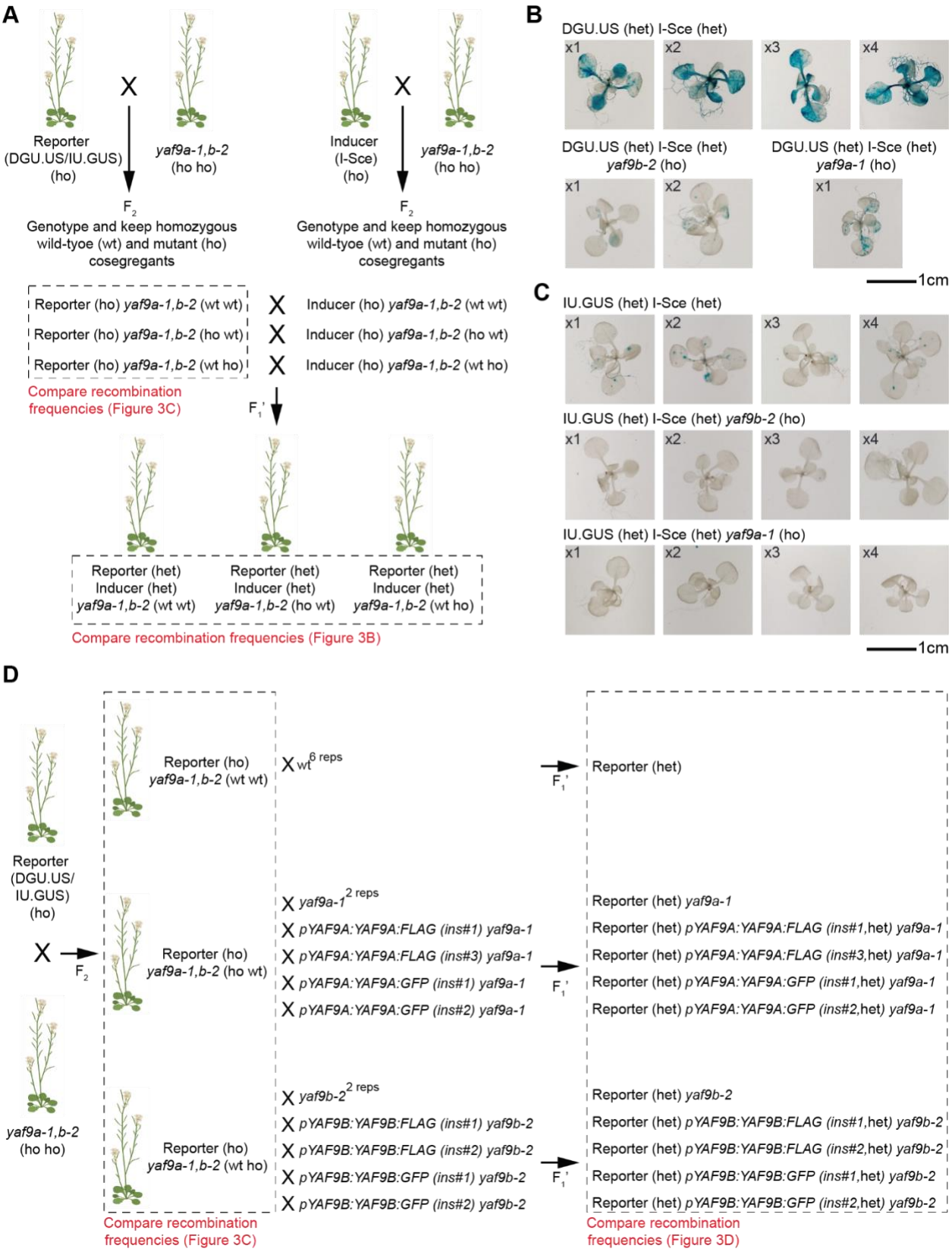

A
Col-0
yaf9b-3
yaf9b-2
yaf9a-1
yaf9a-1,b-3
yaf9a-1,b-2
B
YAF9B
yaf9b-3
yaf9b-2
3654
3655
YEATS domain
wt
+ γ-IR
- γ-IR
yaf9a-1,b-3
+ γ-IR
- γ-IR
50
0
500 bp
C
- γ-IR 3h
+ γ-IR 3h
Col-0
yaf9a−1
yaf9b−3
yaf9b−2
yaf9a−1,b−3
yaf9a−1,b−2
YAF9B
807bp
Act2
228bp
D
YAF9B
YAF9A
yaf9b-3
yaf9b-2
yaf9a-1
YEATS
YEATS
30
0
Col-0
+ γ-IR
yaf9a-1,b-3
+ γ-IR
yaf9a-1,b-2
+ γ-IR
E
No Treatment
75Gy
100Gy
125Gy
% seedlings
100
75
50
25
0
Col-0
yaf9a−1
yaf9b−3
yaf9b−2
yaf9a−1,b−3
yaf9a−1,b−2
Replicate 3
F
pYAF9A:YAF9A:FLAG or GFP
T1 selection
T2/T3 Genotyping
yaf9a-1(ho)
b-2(het)
pYAF9A:YAF9A:FLAG (ins#1)
pYAF9A:YAF9A:GFP (ins#1)
yaf9a−1
yaf9a−1,b−2
pYAF9B:YAF9B:FLAG or GFP
T1 selection
T2/T3 Genotyping
yaf9a-1(het)
b-2(ho)
pYAF9B:YAF9B:FLAG (ins#1)
pYAF9B:YAF9B:GFP (ins#1)
yaf9b-2
yaf9a−1,b−2
5 cm

Figure 3-S3. YAF9A and YAF9B are required for homologous recombination after targeted DSB damage.

(A) Crossing scheme to introgress reporter transgenes (DGU.US or IU.GUS) and inducer transgene (I-Sce), into *yaf9* mutant backgrounds to assess HR-like repair efficiencies. The reporter lines (DGU.US or IU.GUS) as well as an I-SceI inducer line were independently crossed with *yaf9a-1,b-2* double mutant and, in subsequent generations, cosegregants homozygous or wild-type for *yaf9* single/ double mutants were identified by PCR-based genotyping. Finally, the reporters and inducers of the same *yaf9* background were crossed together to generate plants that are heterozygous for both the reporter and the inducer construct in a homozygous mutant or wild-type background. Images showing the somatic recombination events, represented by GUS activity, via **(B)** SSA and **(C)** SDSA in *yaf9* single mutants in reporter (DGU.US or IU,GUS) - inducer (I-Sce) crosses. Each image is a biological replicate originating from a different cross with unique parents. The cross number is indicated on the top-left corner of the image. The scale bar represents 1 cm. **(D)** Crossing scheme to rescue HR-like defects in *yaf9* mutants using complementing lines. The reporter lines (DGU.US or IU.GUS) were independently crossed with *yaf9a-1,b-2* double mutant and, in subsequent generations, cosegregants homozygous or wild-type for *yaf9* single were identified by PCR-based genotyping. The wild-type cosegregants was crossed to a wild-type plant to generate Reporter (het) *YAF9* (wt). The mutant cosegregants was either crossed to *yaf9* mutant plants or independent insertions of FLAG/ GFP tagged complementing lines to generate Reporter (het) *yaf9* (ho) and Reporter (het) *yaf9* (ho) *pYAF9:YAF9:FLAG/GFP* (het) respectively. Plants from which recombination *rates* are compared are highlighted in dashed rectangles.


A
Reporter (DGU.US/IU.GUS) (ho)
X
yaf9a-1,b-2 (ho ho)
F2
Genotype and keep homozygous wild-tyoe (wt) and mutant (ho) cosegregants
Inducer (I-Sce) (ho)
X
yaf9a-1,b-2 (ho ho)
F2
Genotype and keep homozygous wild-tyoe (wt) and mutant (ho) cosegregants
Reporter (ho) yaf9a-1,b-2 (wt wt) X Inducer (ho) yaf9a-1,b-2 (wt wt)
Reporter (ho) yaf9a-1,b-2 (ho wt) X Inducer (ho) yaf9a-1,b-2 (ho wt)
Reporter (ho) yaf9a-1,b-2 (wt ho) X Inducer (ho) yaf9a-1,b-2 (wt ho)
Compare recombination frequencies (Figure 3C)
F1'
Reporter (het) Inducer (het) yaf9a-1,b-2 (wt wt)
Reporter (het) Inducer (het) yaf9a-1,b-2 (ho wt)
Reporter (het) Inducer (het) yaf9a-1,b-2 (wt ho)
Compare recombination frequencies (Figure 3B)
B
DGU.US (het) I-Sce (het)
x1
x2
x3
x4
DGU.US (het) I-Sce (het) yaf9b-2 (ho)
DGU.US (het) I-Sce (het) yaf9a-1 (ho)
1cm
C
IU.GUS (het) I-Sce (het)
IU.GUS (het) I-Sce (het) yaf9b-2 (ho)
IU.GUS (het) I-Sce (het) yaf9a-1 (ho)
1cm
D
Reporter (DGU.US/ IU.GUS) (ho)
X
F2
yaf9a-1,b-2 (ho ho)
Reporter (ho) yaf9a-1,b-2 (wt wt)
X wt 6 reps
Reporter (ho) yaf9a-1,b-2 (ho wt)
X yaf9a-1 2 reps
X pYAF9A:YAF9A:FLAG (ins#1) yaf9a-1
X pYAF9A:YAF9A:FLAG (ins#3) yaf9a-1
X pYAF9A:YAF9A:GFP (ins#1) yaf9a-1
X pYAF9A:YAF9A:GFP (ins#2) yaf9a-1
Reporter (ho) yaf9a-1,b-2 (wt ho)
X yaf9b-2 2 reps
X pYAF9B:YAF9B:FLAG (ins#1) yaf9b-2
X pYAF9B:YAF9B:FLAG (ins#2) yaf9b-2
X pYAF9B:YAF9B:GFP (ins#1) yaf9b-2
X pYAF9B:YAF9B:GFP (ins#2) yaf9b-2
Compare recombination frequencies (Figure 3C)
F1'
F1'
F1'
Reporter (het)
Reporter (het) yaf9a-1
Reporter (het) pYAF9A:YAF9A:FLAG (ins#1,het) yaf9a-1
Reporter (het) pYAF9A:YAF9A:FLAG (ins#3,het) yaf9a-1
Reporter (het) pYAF9A:YAF9A:GFP (ins#1,het) yaf9a-1
Reporter (het) pYAF9A:YAF9A:GFP (ins#2,het) yaf9a-1
Reporter (het) yaf9b-2
Reporter (het) pYAF9B:YAF9B:FLAG (ins#1,het) yaf9b-2
Reporter (het) pYAF9B:YAF9B:FLAG (ins#2,het) yaf9b-2
Reporter (het) pYAF9B:YAF9B:GFP (ins#1,het) yaf9b-2
Reporter (het) pYAF9B:YAF9B:GFP (ins#2,het) yaf9b-2
Compare recombination frequencies (Figure 3D)

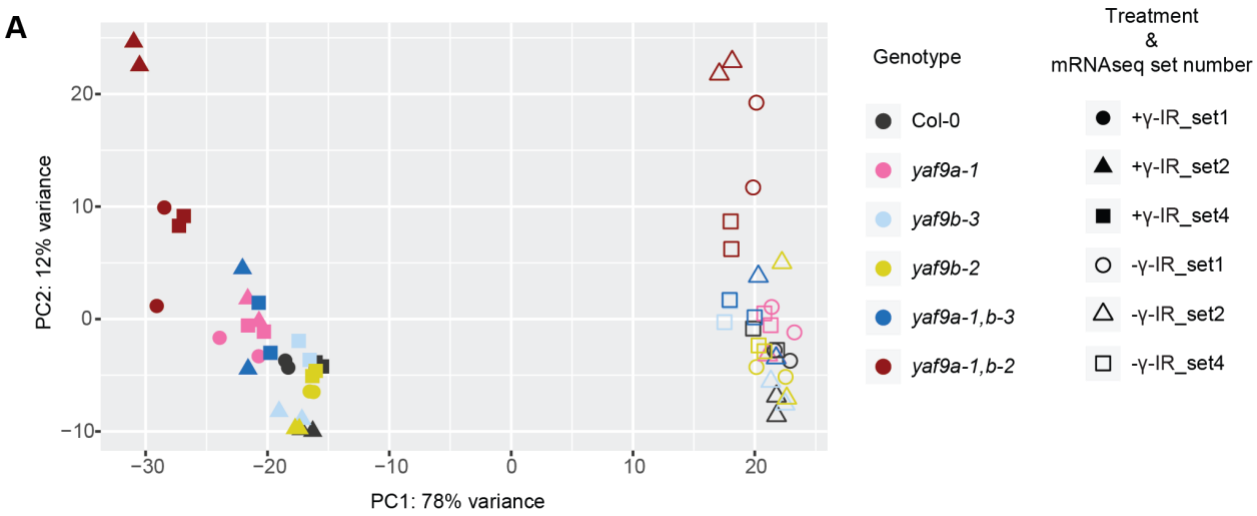


Figure 3-S4. Transcriptional profiles of *yaf9* mutants with and without DNA damage.

(**A**) Principle component analysis (PCA) plots, after batch correction with limma(Ritchie et al., 2015), showing the groupings of data points from the wild-type (Col-0) and *yaf9* single and double mutants 3h post gamma-irradiation (+ γ-IR; 100Gy) or mock (- γ-IR) treatments. Genotypes are shown using matched colors and the two replicates for each sample from 3 different mRNAseq batches (set1, 2 and 4) are represented using matched shapes.

ACKNOWLEDGEMENTS

Chapter 3, in part, is made up of a manuscript in preparation. The manuscript in preparation is: Neeraja Vegesna, Clara Bourbousse, Yasaman Jami-Alahmadi, Ana Marie S. Palanca, James A. Wohlschlegel, Julie A. Law**. "Two *Arabidopsis* YEATS domain proteins facilitate DNA damage repair via their histone binding domains"** *(in preparation)*. The dissertation author will be the first author of this manuscript in preparation.

REFERENCES

Aguilera, A., & Garcia-Muse, T. (2013). Causes of genome instability. *Annu Rev Genet, 47*, 1-32. doi:10.1146/annurev-genet-111212-133232

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., . . . Ecker, J. R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science, 301*(5633), 653-657. doi:10.1126/science.1086391

Andrews, F. H., Shinsky, S. A., Shanle, E. K., Bridgers, J. B., Gest, A., Tsun, I. K., . . . Kutateladze, T. G. (2016). The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol, 12*(6), 396-398. doi:10.1038/nchembio.2065

Aslam, M., Fakher, B., Jakada, B. H., Cao, S., & Qin, Y. (2019). SWR1 Chromatin Remodeling Complex: A Key Transcriptional Regulator in Plants. *Cells, 8*(12). doi:10.3390/cells8121621

Barnes, C. E., English, D. M., & Cowley, S. M. (2019). Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. *Essays Biochem, 63*(1), 97-107. doi:10.1042/EBC20180061

Bieluszewski, T., Galganski, L., Sura, W., Bieluszewska, A., Abram, M., Ludwikow, A., . . . Sadowski, J. (2015). AtEAF1 is a potential platform protein for Arabidopsis NuA4 acetyltransferase complex. *BMC Plant Biol, 15*, 75. doi:10.1186/s12870-015-0461-1

Bohmdorfer, G., Schleiffer, A., Brunmeir, R., Ferscha, S., Nizhynska, V., Kozak, J., . . . Schweizer, D. (2011). GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in Arabidopsis. *Plant J, 67*(3), 420-433. doi:10.1111/j.1365-313X.2011.04604.x

Bourbousse, C., Vegesna, N., & Law, J. A. (2018). SOG1 activator and MYB3R repressors regulate a complex DNA damage network in Arabidopsis. *Proc Natl Acad Sci U S A, 115*(52), E12453-E12462. doi:10.1073/pnas.1810582115

Bouyer, D., Heese, M., Chen, P., Harashima, H., Roudier, F., Gruttner, C., & Schnittger, A. (2018). Genome-wide identification of RETINOBLASTOMA RELATED 1 binding sites in Arabidopsis reveals novel DNA damage regulators. *PLoS Genet, 14*(11), e1007797. doi:10.1371/journal.pgen.1007797

Campi, M., D'Andrea, L., Emiliani, J., & Casati, P. (2012). Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. *Plant Physiol, 158*(2), 981-995. doi:10.1104/pp.111.191452

Ciccia, A., & Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. *Mol Cell, 40*(2), 179-204. doi:10.1016/j.molcel.2010.09.019

Clouaire, T., Rocher, V., Lashgari, A., Arnould, C., Aguirrebengoa, M., Biernacka, A., . . . Legube, G. (2018). Comprehensive Mapping of Histone Modifications at DNA Double-Strand

Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell, 72*(2), 250-262 e256. doi:10.1016/j.molcel.2018.08.020

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J, 16*(6), 735-743. doi:10.1046/j.1365-313x.1998.00343.x

Crevillen, P., Gomez-Zambrano, A., Lopez, J. A., Vazquez, J., Pineiro, M., & Jarillo, J. A. (2019). Arabidopsis YAF9 histone readers modulate flowering time through NuA4-complex-dependent H4 and H2A.Z histone acetylation at FLC chromatin. *New Phytol, 222*(4), 1893-1908. doi:10.1111/nph.15737

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Drury, G. E., Dowle, A. A., Ashford, D. A., Waterworth, W. M., Thomas, J., & West, C. E. (2012). Dynamics of plant histone modifications in response to DNA damage. *Biochem J, 445*(3), 393-401. doi:10.1042/BJ20111956

Espinosa-Cores, L., Bouza-Morcillo, L., Barrero-Gil, J., Jimenez-Suarez, V., Lazaro, A., Piqueras, R., . . . Pineiro, M. (2020). Insights Into the Function of the NuA4 Complex in Plants. *Front Plant Sci, 11*, 125. doi:10.3389/fpls.2020.00125

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., & Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. *DNA Repair (Amst), 3*(8-9), 959-967. doi:10.1016/j.dnarep.2004.03.024

Fina, J. P., & Casati, P. (2015). HAG3, a Histone Acetyltransferase, Affects UV-B Responses by Negatively Regulating the Expression of DNA Repair Enzymes and Sunscreen Content in Arabidopsis thaliana. *Plant Cell Physiol, 56*(7), 1388-1400. doi:10.1093/pcp/pcv054

Fina, J. P., Masotti, F., Rius, S. P., Crevacuore, F., & Casati, P. (2017). HAC1 and HAF1 Histone Acetyltransferases Have Different Roles in UV-B Responses in Arabidopsis. *Front Plant Sci, 8*, 1179. doi:10.3389/fpls.2017.01179

Friesner, J. D., Liu, B., Culligan, K., & Britt, A. B. (2005). Ionizing radiation-dependent gamma-H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. *Mol Biol Cell, 16*(5), 2566-2576. doi:10.1091/mbc.e04-10-0890

Georgoulis, A., Vorgias, C. E., Chrousos, G. P., & Rogakou, E. P. (2017). Genome Instability and gammaH2AX. *Int J Mol Sci, 18*(9). doi:10.3390/ijms18091979

Gomez, M. S., Falcone Ferreyra, M. L., Sheridan, M. L., & Casati, P. (2019). Arabidopsis E2Fc is required for the DNA damage response under UV-B radiation epistatically over the microRNA396 and independently of E2Fe. *Plant J, 97*(4), 749-764. doi:10.1111/tpj.14158

Gong, F., Chiu, L. Y., & Miller, K. M. (2016). Acetylation Reader Proteins: Linking Acetylation Signaling to Genome Maintenance and Cancer. *PLoS Genet, 12*(9), e1006272. doi:10.1371/journal.pgen.1006272

Goudarzi, A., Zhang, D., Huang, H., Barral, S., Kwon, O. K., Qi, S., . . . Khochbin, S. (2016). Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters. *Mol Cell, 62*(2), 169-180. doi:10.1016/j.molcel.2016.03.014

Hafner, A., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nat Rev Mol Cell Biol, 20*(4), 199-210. doi:10.1038/s41580-019-0110-x

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell, 38*(4), 576-589. doi:10.1016/j.molcel.2010.05.004

Hirakawa, T., Hasegawa, J., White, C. I., & Matsunaga, S. (2017). RAD54 forms DNA repair foci in response to DNA damage in living plant cells. *Plant J, 90*(2), 372-382. doi:10.1111/tpj.13499

Hong, J. H., Savina, M., Du, J., Devendran, A., Kannivadi Ramakanth, K., Tian, X., . . . Xu, J. (2017). A Sacrifice-for-Survival Mechanism Protects Root Stem Cell Niche from Chilling Stress. *Cell, 170*(1), 102-113 e114. doi:10.1016/j.cell.2017.06.002

Hsu, C. C., Zhao, D., Shi, J., Peng, D., Guan, H., Li, Y., . . . Shi, X. (2018). Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov, 4*(1), 28. doi:10.1038/s41421-018-0027-0

Hu, Z., Cools, T., & De Veylder, L. (2016). Mechanisms Used by Plants to Cope with DNA Damage. *Annu Rev Plant Biol, 67*, 439-462. doi:10.1146/annurev-arplant-043015-111902

Jeggo, P. A., Downs, J. A., & Gasser, S. M. (2017). Chromatin modifiers and remodellers in DNA repair and signalling. *Philos Trans R Soc Lond B Biol Sci, 372*(1731). doi:10.1098/rstb.2016.0279

Kadoch, C., & Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv, 1*(5), e1500447. doi:10.1126/sciadv.1500447

Karimi, M., Depicker, A., & Hilson, P. (2007). Recombinational cloning with plant gateway vectors. *Plant Physiol, 145*(4), 1144-1154. doi:10.1104/pp.107.106989

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. https://CRAN.R-project.org/package=ggpubr

Kebede, A. F., Nieborak, A., Shahidian, L. Z., Le Gras, S., Richter, F., Gomez, D. A., . . . Schneider, R. (2017). Histone propionylation is a mark of active chromatin. *Nat Struct Mol Biol, 24*(12), 1048-1056. doi:10.1038/nsmb.3490

Kevin Blighe, S. R., Myles Lewis. (2021). Publication-ready volcano plots with enhanced colouring and labeling. doi:https://doi.org/doi:10.18129/B9.bioc.EnhancedVolcano

Klein, B. J., Ahmad, S., Vann, K. R., Andrews, F. H., Mayo, Z. A., Bourriquen, G., . . . Kutateladze, T. G. (2018). Yaf9 subunit of the NuA4 and SWR1 complexes targets histone H3K27ac through its YEATS domain. *Nucleic Acids Res, 46*(1), 421-430. doi:10.1093/nar/gkx1151

Klein, B. J., Vann, K. R., Andrews, F. H., Wang, W. W., Zhang, J., Zhang, Y., . . . Kutateladze, T. G. (2018). Structural insights into the pi-pi-pi stacking mechanism and DNA-binding activity of the YEATS domain. *Nat Commun, 9*(1), 4574. doi:10.1038/s41467-018-07072-6

Lario, L. D., Ramirez-Parra, E., Gutierrez, C., Casati, P., & Spampinato, C. P. (2011). Regulation of plant MSH2 and MSH6 genes in the UV-B-induced DNA damage response. *J Exp Bot, 62*(8), 2925-2937. doi:10.1093/jxb/err001

Law, J. A., Ausin, I., Johnson, L. M., Vashisht, A. A., Zhu, J. K., Wohlschlegel, J. A., & Jacobsen, S. E. (2010). A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. *Curr Biol, 20*(10), 951-956. doi:10.1016/j.cub.2010.03.062

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, Y., Sabari, B. R., Panchenko, T., Wen, H., Zhao, D., Guan, H., . . . Li, H. (2016). Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell, 62*(2), 181-193. doi:10.1016/j.molcel.2016.03.028

Li, Y., Zhao, D., Chen, Z., & Li, H. (2017). YEATS domain: Linking histone crotonylation to gene regulation. *Transcription, 8*(1), 9-14. doi:10.1080/21541264.2016.1239602

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lu, P. Y., Levesque, N., & Kobor, M. S. (2009). NuA4 and SWR1-C: two chromatin-modifying complexes with overlapping functions and components. *Biochem Cell Biol, 87*(5), 799-815. doi:10.1139/O09-062

Luo, M., Cheng, K., Xu, Y., Yang, S., & Wu, K. (2017). Plant Responses to Abiotic Stress Regulated by Histone Deacetylases. *Front Plant Sci, 8*, 2147. doi:10.3389/fpls.2017.02147

Luo, Y. X., Hou, X. M., Zhang, C. J., Tan, L. M., Shao, C. R., Lin, R. N., . . . He, X. J. (2020). A plant-specific SWR1 chromatin-remodeling complex couples histone H2A.Z deposition with nucleosome sliding. *EMBO J, 39*(7), e102008. doi:10.15252/embj.2019102008

Maehara, K., Harada, A., Sato, Y., Matsumoto, M., Nakayama, K. I., Kimura, H., & Ohkawa, Y. (2015). Tissue-specific expression of histone H3 variants diversified after species separation. *Epigenetics Chromatin, 8*, 35. doi:10.1186/s13072-015-0027-3

Mahapatra, K., & Roy, S. (2021). SOG1 transcription factor promotes the onset of endoreduplication under salinity stress in Arabidopsis. *Sci Rep, 11*(1), 11659. doi:10.1038/s41598-021-91293-1

Manova, V., & Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. *Front Plant Sci, 6*, 885. doi:10.3389/fpls.2015.00885

Mintoff, S. J., Rookes, J. E., & Cahill, D. M. (2015). Sub-lethal UV-C radiation induces callose, hydrogen peroxide and defence-related gene expression in Arabidopsis thaliana. *Plant Biol (Stuttg), 17*(3), 703-711. doi:10.1111/plb.12286

Nguyen, V. Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., & Leschziner, A. E. (2013). Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell, 154*(6), 1220-1231. doi:10.1016/j.cell.2013.08.018

Nitarska, J., Smith, J. G., Sherlock, W. T., Hillege, M. M., Nott, A., Barshop, W. D., . . . Riccio, A. (2016). A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Rep, 17*(6), 1683-1698. doi:10.1016/j.celrep.2016.10.022

Ogita, N., Okushima, Y., Tokizawa, M., Yamamoto, Y. Y., Tanaka, M., Seki, M., . . . Umeda, M. (2018). Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in Arabidopsis. *Plant J, 94*(3), 439-453. doi:10.1111/tpj.13866

Poklar, N., Pilch, D. S., Lippard, S. J., Redding, E. A., Dunham, S. U., & Breslauer, K. J. (1996). Influence of cisplatin intrastrand crosslinking on the conformation, thermal stability, and energetics of a 20-mer DNA duplex. *Proc Natl Acad Sci U S A, 93*(15), 7606-7611. doi:10.1073/pnas.93.15.7606

Potok, M. E., Wang, Y., Xu, L., Zhong, Z., Liu, W., Feng, S., . . . Jacobsen, S. E. (2019). Arabidopsis SWR1-associated protein methyl-CpG-binding domain 9 is required for histone H2A.Z deposition. *Nat Commun, 10*(1), 3352. doi:10.1038/s41467-019-11291-w

Puchta, H., & Hohn, B. (2012). In planta somatic homologous recombination assay revisited: a successful and versatile, but delicate tool. *Plant Cell, 24*(11), 4324-4331. doi:10.1105/tpc.112.101824

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007

Rodgers, K., & McVey, M. (2016). Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol, 231*(1), 15-24. doi:10.1002/jcp.25053

Rosa, M., & Scheid, O. M. (2014a). DNA Damage Sensitivity Assays with Arabidopsis Seedlings. *Bio Protoc, 4*(7). doi:10.21769/BioProtoc.1093

Rosa, M., & Scheid, O. M. (2014b). Measuring Homologous Recombination Frequency in Arabidopsis Seedlings. *Bio Protoc, 4*(7). doi:10.21769/BioProtoc.1094

Rosa, M., Von Harder, M., Cigliano, R. A., Schlogelhofer, P., & Mittelsten Scheid, O. (2013). The Arabidopsis SWR1 chromatin-remodeling complex is important for DNA repair, somatic recombination, and meiosis. *Plant Cell, 25*(6), 1990-2001. doi:10.1105/tpc.112.104067

Roth, N., Klimesch, J., Dukowic-Schulze, S., Pacher, M., Mannuss, A., & Puchta, H. (2012). The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. *Plant J, 72*(5), 781-790. doi:10.1111/j.1365-313X.2012.05119.x

Ryu, T. H., Go, Y. S., Choi, S. H., Kim, J. I., Chung, B. Y., & Kim, J. H. (2019). SOG1-dependent NAC103 modulates the DNA damage response as a transcriptional regulator in Arabidopsis. *Plant J, 98*(1), 83-96. doi:10.1111/tpj.14201

Saban, N., & Bujak, M. (2009). Hydroxyurea and hydroxamic acid derivatives as antitumor drugs. *Cancer Chemother Pharmacol, 64*(2), 213-221. doi:10.1007/s00280-009-0991-z

Sabari, B. R., Tang, Z., Huang, H., Yong-Gonzalez, V., Molina, H., Kong, H. E., . . . Allis, C. D. (2015). Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol Cell, 58*(2), 203-215. doi:10.1016/j.molcel.2015.02.029

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci U S A, 103*(49), 18822-18827. doi:10.1073/pnas.0605639103

Schulze, J. M., Wang, A. Y., & Kobor, M. S. (2009). YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochem Cell Biol, 87*(1), 65-75. doi:10.1139/O08-111

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol, 20*(11), 698-714. doi:10.1038/s41580-019-0152-0

Seo, S., Maeda, T., & Hiratsuka, K. (2007). Tissue-specific and DNA damage-responsive expression of the Arabidopsis RAD51 gene promoter in transgenic Arabidopsis and tobacco. *Plant Biotechnology, 24*(3), 321-329. doi:10.5511/plantbiotechnology.24.321

Setiaputra, D., Ahmad, S., Dalwadi, U., Steunou, A. L., Lu, S., Ross, J. D., . . . Yip, C. K. (2018). Molecular Architecture of the Essential Yeast Histone Acetyltransferase Complex NuA4 Redefines Its Multimodularity. *Mol Cell Biol, 38*(9). doi:10.1128/MCB.00570-17

Siddiqui, M. S., Francois, M., Fenech, M. F., & Leifert, W. R. (2015). Persistent gammaH2AX: A promising molecular marker of DNA damage and aging. *Mutat Res Rev Mutat Res, 766*, 1-19. doi:10.1016/j.mrrev.2015.07.001

Sijacic, P., Holder, D. H., Bajic, M., & Deal, R. B. (2019). Methyl-CpG-binding domain 9 (MBD9) is required for H2A.Z incorporation into chromatin at a subset of H2A.Z-enriched regions in the Arabidopsis genome. *PLoS Genet, 15*(8), e1008326. doi:10.1371/journal.pgen.1008326

Squatrito, M., Gorrini, C., & Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. *Trends Cell Biol, 16*(9), 433-442. doi:10.1016/j.tcb.2006.07.007

Srivastava, N., Gochhait, S., de Boer, P., & Bamezai, R. N. K. (2009). Role of H2AX in DNA damage response and human cancers. *Mutat Res, 681*(2-3), 180-188. doi:10.1016/j.mrrev.2008.08.003

Takahashi, N., Ogita, N., Takahashi, T., Taniguchi, S., Tanaka, M., Seki, M., & Umeda, M. (2019). A regulatory module controlling stress-induced cell cycle arrest in Arabidopsis. *Elife, 8*. doi:10.7554/eLife.43944

Tan, L. M., Zhang, C. J., Hou, X. M., Shao, C. R., Lu, Y. J., Zhou, J. X., . . . He, X. J. (2018). The PEAT protein complexes are required for histone deacetylation and heterochromatin silencing. *EMBO J, 37*(19). doi:10.15252/embj.201798770

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res, 22*(22), 4673-4680. doi:10.1093/nar/22.22.4673

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform, 14*(2), 178-192. doi:10.1093/bib/bbs017

Ulker, B., Hommelsheim, C. M., Berson, T., Thomas, S., Chandrasekar, B., Olcay, A. C., . . . Frantzeskakis, L. (2012). Reevaluation of the reliability and usefulness of the somatic homologous recombination reporter lines. *Plant Cell, 24*(11), 4314-4323. doi:10.1105/tpc.112.100404

Van, H. T., & Santos, M. A. (2018). Histone modifications and the DNA double-strand break response. *Cell Cycle, 17*(21-22), 2399-2410. doi:10.1080/15384101.2018.1542899

Wang, L., Ma, K. B., Lu, Z. G., Ren, S. X., Jiang, H. R., Cui, J. W., . . . Jin, B. (2020). Differential physiological, transcriptomic and metabolomic responses of Arabidopsis leaves under prolonged warming and heat shock. *BMC Plant Biol, 20*(1), 86. doi:10.1186/s12870-020-2292-y

Wang, Y., Jin, J., Chung, M. W. H., Feng, L., Sun, H., & Hao, Q. (2018). Identification of the YEATS domain of GAS41 as a pH-dependent reader of histone succinylation. *Proc Natl Acad Sci U S A, 115*(10), 2365-2370. doi:10.1073/pnas.1717664115

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *Plos One, 2*(8), e718. doi:10.1371/journal.pone.0000718

Woody, S. T., Austin-Phillips, S., Amasino, R. M., & Krysan, P. J. (2007). The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res, 120*(1), 157-165. doi:10.1007/s10265-006-0048-x

Yoshiyama, K., Conklin, P. A., Huefner, N. D., & Britt, A. B. (2009). Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A, 106*(31), 12843-12848. doi:10.1073/pnas.0810304106

Yoshiyama, K. O. (2016). SOG1: a master regulator of the DNA damage response in plants. *Genes Genet Syst, 90*(4), 209-216. doi:10.1266/ggs.15-00011

Yoshiyama, K. O., Kobayashi, J., Ogita, N., Ueda, M., Kimura, S., Maki, H., & Umeda, M. (2013). ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in Arabidopsis. *EMBO Rep, 14*(9), 817-822. doi:10.1038/embor.2013.112

Yu, H., Kong, X., Huang, H., Wu, W., Park, J., Yun, D. J., . . . Zhu, J. K. (2020). STCH4/REIL2 Confers Cold Stress Tolerance in Arabidopsis by Promoting rRNA Processing and CBF Protein Translation. *Cell Rep, 30*(1), 229-242 e225. doi:10.1016/j.celrep.2019.12.012

Zacharaki, V., Benhamed, M., Poulios, S., Latrasse, D., Papoutsoglou, P., Delarue, M., & Vlachonasios, K. E. (2012). The Arabidopsis ortholog of the YEATS domain containing protein YAF9a regulates flowering by controlling H4 acetylation levels at the FLC locus. *Plant Sci, 196*, 44-52. doi:10.1016/j.plantsci.2012.07.010

Zhang, Q., Zeng, L., Zhao, C., Ju, Y., Konuma, T., & Zhou, M. M. (2016). Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure, 24*(9), 1606-1612. doi:10.1016/j.str.2016.05.023

Zhao, D., Guan, H., Zhao, S., Mi, W., Wen, H., Li, Y., . . . Li, H. (2016). YEATS2 is a selective histone crotonylation reader. *Cell Res, 26*(5), 629-632. doi:10.1038/cr.2016.49

Zhao, D., Li, Y., Xiong, X., Chen, Z., & Li, H. (2017). YEATS Domain-A Histone Acylation Reader in Health and Disease. *J Mol Biol*, *429*(13), 1994-2002. doi:10.1016/j.jmb.2017.03.010

Zhao, S., Zhang, B., Yang, M., Zhu, J., & Li, H. (2018). Systematic Profiling of Histone Readers in Arabidopsis thaliana. *Cell Rep*, *22*(4), 1090-1102. doi:10.1016/j.celrep.2017.12.099

Zhou, M., Coruh, C., Xu, G., Bourbousse, C., Lambolez, A., & Law, J. A. (2021). The CLASSY family controls tissue-specific DNA methylation patterns in Arabidopsis. *bioRxiv*, 2021.2001.2023.427869. doi:10.1101/2021.01.23.427869

SUPPLEMENTARY INFORMATION

Supplementary table 1: Summary of mRNA-seq data and mapping

Supplementary table 2: (A) Expression values of genes in yaf9 mRNAseq, (B) list of differentially expressed genes and (C) log2FC and adjusted p-values from DESeq.

Supplementary table 3: List of primers used.

Supplementary table 4: Details of cloning constructs: Gateway cloning BP reactions and resultant entry clones; LR reactions and resultant expression clones

Supplementary table 5: Protein enrichment values in YAF9B: FLAG and Col-0 IP/MS, 6h after 100Gy.

Chapter 4 CONCLUSION AND FUTURE DIRECTIONS

This work applied a genetics, genomics, and biochemical approaches to expand our understanding of two central unknowns of the *Arabidopsis* DNA damage response (DDR).

The first is the transcriptional regulation of the DDR and its downstream processes. To address this gap, we generated the first model of the DNA damage response transcriptional network, revealing 11 co-expressed gene groups with distinct biological functions and cis-regulatory features (Chapter 2)(Bourbousse, Vegesna, & Law, 2018). In addition, our characterization of this model demonstrates that SOG1 and three MYB3R TFs are, respectively, the major activator and repressors within this network, coordinating the rapid induction of DNA repair genes and TF cascades, as well as the subsequent repression of cell cycle genes. This time-course transcriptional model is the most comprehensive map of the *Arabidopsis* DDR so far and hence, serves as a great resource for other scientists to interpret co-expression patterns for their genes of interest. Moreover, it serves as a framework for understanding the complex DNA damage response and the insights from it could pave way to the generation of new hypotheses. For instance, we have already used co-expression insights from this model to set up a reverse genetic screen (Chapter 3). Further characterization of the SOG1-independent DDR, revealed in this model, using additional genetic perturbations and transcriptomics, would perhaps identify novel DDR regulators. In addition to transcriptional regulation, the DDR consists of signaling pathways that are mediated by several post translational modifications. Perhaps, further proteomic experiments exploring post-translational changes after DNA damage would decipher an additional layer of regulation and greatly contribute to a deeper understanding of DDR. Data analysis is another area where there is great potential for additional characterization. While we have identified

differentially expressed genes, the current genome annotation (TAIR10 and Araport11) of *Arabidopsis*(Cheng et al., 2017) does not include a comprehensive list of ncRNAs. As I outlined in the introduction, in mammals, lncRNAs participate in DNA repair, via diverse mechanisms. The roles of lncRNA in DNA repair remains almost unexplored in *Arabidopsis*. Identification and characterization of even a small portion of these lncRNAs in DDR would represent a massive step forward in exploring the "dark matter" of the genome in DDR.

The second unknown is the role of chromatin in orchestrating DDR. To address this gap, we set up a reverse genetic screen to identify chromatin effectors in DDR that was based on the *Arabidopsis* gene co-expression model after DNA damage, (Chapter 3). This strategy led to the identification of 53 candidate genes, including 4 non-coding RNAs, 9 putative repair factors, 11 putative chromatin-associated factors, 29 transcription factors and coactivators. We have already functionally characterized one chromatin-associated factor, YAF9B and connected it to repair. Using various expression profiling experiments, DNA repair assays, and biochemical experiments, the *Arabidopsis* YEATS domain-containing proteins, YAF9A and YAF9B, were both shown to be required for proper DSB-repair via HR-like pathways. Moreover, the characterization showed that the role of YAF9B in DSB repair is chromatin-mediated and dependent on its YEATS domain. This work establishes YAF9B (and YAF9A) as only the second histone readers characterized in DSB repair, and more broadly, connects chromatin to DDR. However, few questions regarding the role of YAF9B still remain. Firstly, if it is acting at the site of DNA damage, which could be confirmed by examining co-localization with DSB-marker γH2A.X. Secondly, if YAF9B's participation in the NuA4 histone acetylation complex results in DSB- specific or global chromatin changes. This could be examined via profiling the H4.ac and H2A.Zac marks around the DSB and globally.

Altogether, while the reverse genetic screening approach is biased to identification factors that are differentially expressed after damage, characterization of novel factors such as, YAF9B in chromatin-mediated aspects of the DDR is a proof of concept that this strategy we pursued works. Therefore, pursuing the other candidate genes selected using the screening approach, would provide additional insights into chromatin mediation of DDR pathways. In addition to conserved aspects of the DDR, *Arabidopsis* is a great model to study plant-specific novel DDR mechanisms. Since the current understanding of the role of chromatin in DDR remains poor in *Arabidopsis*, setting up additional unbiased forward genetic screens, leveraging molecular repair phenotypes, to identify chromatin factors, is important to understand repair.

REFERENCES:

Bourbousse, C., Vegesna, N., & Law, J. A. (2018). SOG1 activator and MYB3R repressors regulate a complex DNA damage network in Arabidopsis. *Proc Natl Acad Sci U S A, 115*(52), E12453-E12462. doi:10.1073/pnas.1810582115

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J, 89*(4), 789-804. doi:10.1111/tpj.13415