U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
ANO

U.PORTO

Gene Regulation at the Chromatin Level in Arabidopsis: DNA-RNA hybrids formation at *DOG1* promoter controls sense and antisense expression

Miguel Montez Coelho

FC

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Gene Regulation at the Chromatin Level in Arabidopsis: DNA-RNA hybrids formation at *DOG1* promoter controls sense and antisense expression

Miguel Montez Coelho

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Biologia Funcional e Biotecnologia de Plantas

2018

MSc

Gene Regulation at the Chromatin Level in Arabidopsis: DNA-RNA hybrids formation at *DOG1* promoter controls sense and antisense expression

Miguel Montez Coelho
Biologia Funcional e Biotecnologia de Plantas
Institute of Biochemistry and Biophysics PAS/Departamento de Biologia FCUP
2018

Orientador
Szymon Swiezewski, PhD, Institute of Biochemistry and Biophysics PAS

Coorientador
Jorge Teixeira, Professor Auxiliar, FCUP

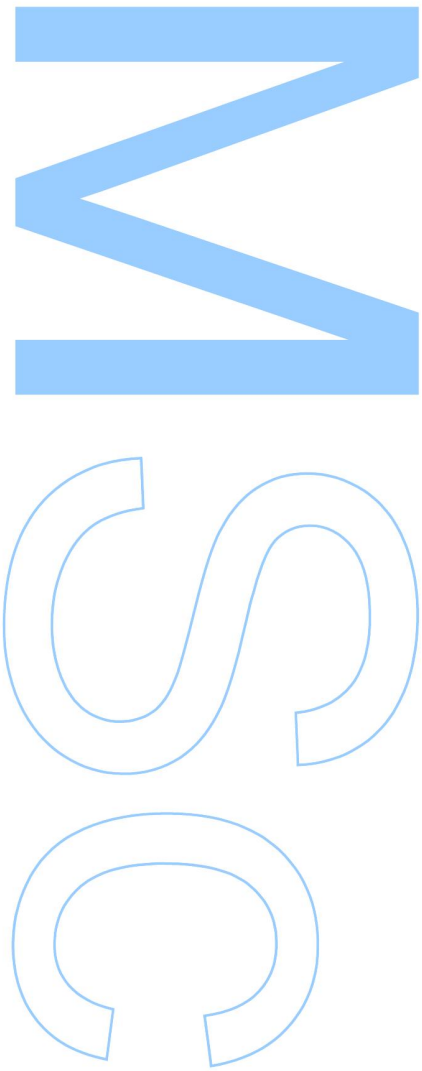
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Acknowledgments

It is with great satisfaction that I write these words to express my true feelings and acknowledge everyone who helped me during this last year.

I want to start by thanking my supervisor Szymon Swiezewski for being present and helping me every time I needed. He started helping me even before my arrival in Warsaw and since then he has giving his support in my every step in the lab and encouraging me not to work too much and live my life outside the lab too…

I must also express my gratitude for my supervisor Prof Jorge Teixeira who presented me a field that fascinated me and which I am fully devoted to contribute to with my research. Even in a long-distance supervision Prof Jorge Teixeira was always present to answer my emails and chat through Skype helping me more than I asked for.

The environment in the lab is of course very important and so I thank very kindly to all the people who contributed to it. When I first arrived to the lab I was warmly welcomed by a few people that are now more than colleagues. Paulina, Halina, Julia, Tomasz, Justyna, Kasia, Lien, Ruslan, Grzesiek and Kris. Later a few more people joined to my daily life. Ferran, Umesh, Wang Ce, Ola, Michal and Sebastian. A special thanks to you all.

At last I express my eternal words of gratitude for my mother and my father who I love for allowing me to follow my dreams. A big and warm hug to my brother, grandma and Vera that are always there for me.

To all my family and friends, thank you for everything!

Resumo

Em eucariotas superiores, todas as células vivas contêm informação genética idêntica, ainda assim executam diferentes programas de expressão génica que estão na base do crescimento, diferenciação, desenvolvimento e resposta a fatores ambientais. A expressão génica é regulada por uma admirável multitude de mecanismos que atuam a vários níveis. Os "R-loops" consistem em híbridos de DNA-RNA em conjunto com uma cadeia simples de DNA que permanece livre. Estas estruturas são formadas principalmente aquando do processo de transcrição e replicação do DNA e são abundantes nos genomas de mamíferos, leveduras e plantas. Nos últimos anos, os "R-loops" têm vindo a ser considerados uma classe importante de elementos reguladores da expressão génica. A formação indesejada de "R-loops" em humanos foi associada com o espectro do autismo e múltiplas doenças neurodegenerativas. Ainda assim, em plantas, o impacto destes elementos regulatórios na expressão génica e integridade do genoma permanecem pouco elucidados. Em Arabidopsis existe apenas um estudo referente a um "R-loop" nuclear funcional formado onde o RNA *antisense* não codificante se encontra no *locus FLC*, integrando a sinalização de baixas temperaturas na regulação da expressão génica no *locus FLC*, que resulta no controlo da floração. Recentemente um estudo a nível global do genoma mapeou os "R-loops" em Arabidopsis, revelando a abundância destas estruturas e a relação com outros elementos regulatórios. Neste estudo foi investigada a formação de "R-loops" no *locus DOG1*. *DOG1* é o principal gene no controlo da dormência de sementes em Arabidopsis. Para além deste papel fundamental no ciclo de vida das plantas, *DOG1* também tem um papel importante em plantas adultas sujeitas a secura. O *locus DOG1* é regulado pelo seu transcrito *antisense asDOG1* entre muitos outros fatores. *asDOG1* é um RNA não codificante que atua em *cis* reprimindo *DOG1* que, por sua vez, também reprime *asDOG1*. Neste estudo foi utilizada a técnica de imunoprecipitação de híbridos de DNA-RNA (DRIP) para mapear um "R-loop" na região do promotor do gene *DOG1*. Foi sugerido que a formação de "R-loops" regula a transcrição *sense* e *antisense* no *locus DOG1*. Foi proposto um modelo onde os "R-loops" regulam eventos de iniciação da transcrição resultando na coordenação do processo transcricional que previne a repressão mútua entre *DOG1 sense* e *antisense* e conduz à ótima produção de transcritos *sense* e *antisense* ao nível de um *locus* individual.

Abstract

In higher eukaryotes all living cells contain identical genetic information, yet they can execute different gene expression programs that underlay cell growth, differentiation, development and responses to environmental cues. Gene expression is regulated by a remarkable multitude of mechanisms that act at various layers. R-loops consist of a DNA-RNA hybrid and a displaced single strand of DNA. These structures are formed mainly during transcription and replication, and are abundant throughout the genome in mammals, yeast and plants. In the recent years, R-loops have been considered to constitute an important class of regulators of gene expression. Unscheduled R-loops formation in humans is known to be associated with autism-spectrum disorders and multiple neurodegenerative diseases. Yet in plants, the impact of this regulatory elements on gene expression and genome integrity remains poorly elucidated. In Arabidopsis only one example of a functional nuclear R-loop was reported, at the antisense lncRNA *COOLAIR* at *FLC locus* integrating the cold sensing in the regulation of *FLC* expression, and ultimately controlling the flowering time. Recently, a genome-wide study provided the mapping of R-loops throughout the Arabidopsis genome, revealing their vast abundancy and links with other regulatory elements. In this study, R-loops formation was investigated within the *DOG1 locus*. *DOG1* is the master regulatory gene controlling seed dormancy in Arabidopsis. In addition to its essential role early in the plant's life cycle, *DOG1* is also a player in the response to drought stress in adult plants. *DOG1* is regulated by its antisense transcript, *asDOG1*, among many other factors. *asDOG1* is a *cis*-acting lncRNA that regulates *DOG1* in a negative feedback loop. Here, DNA-RNA Immunoprecipitation (DRIP) was used to map a R-loop at *DOG1* promoter. R-loops formation was suggested to regulate both sense and antisense transcription at *DOG1 locus*. A model is proposed in which R-loops regulate transcription bursts resulting in coordinated transcription that prevents the sense-antisense mutual repression and leads to optimal sense and antisense transcription status at a single *locus*.

Key-words

Arabidopsis thaliana, gene expression, R-loops, *DOG1*, lncRNA, sense, antisense, DNA-RNA Immunoprecipitation, transcription bursts

Table of Contents

Table Index

Figure Index

region (black lines); exon 2 in the *shDOG1* is extended (white box); exonic regions derived from alternative splicing (grey boxes) (adapted from Fedak et al., 2016).

Fig. 5 – **CPT effect on sense and antisense *DOG1* expression**. RT-qPCR for *DOG1*, *asDOG1* (*At5G45830.1*) and *GP1* (positive control; Dinh et al., 2014) in Col-0 seedlings growing in the presence of 2 μM of CPT 25 μM of CPT or DMSO as control. Expression levels were normalized against *UBC21* (*At5G25760.1*) mRNA; data represents the means of three biological replicates for each treatment condition with error bars representing standard deviation. * show significant differences with *t*-test for $p < 0.05$.

Fig. 6 – **Schematic diagrams of the reporter constructs**. Constructs used to obtain the reporter transgenic lines (performed before by members of the laboratory) containing the luciferase reporter cassette fused to *DOG1* in the genomic context (*pDOG1-LUC::DOG1* referred as genSense), fused to *DOG1* in separated promoter (*pDOG1::LUC* referred as pSense), fused to *asDOG1* in separated promoter (*pasDOG1::LUC* referred as pAS) (Fedak et al., 2016), and fused to *asDOG1* in the genomic context (*pasDOG1-LUC::DOG1* referred as pAS) (unpublished). The *asDOG1* constructs contain an additional IRES sequence to drive translation of the RNA transcripts.

Fig. 7 – ***DOG1* expression changes in response to CPT in genSense plants**. Representative picture of 12-days old Col-0 seedlings carrying *psDOG1-LUC::DOG1* (genSense) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *DOG1* expression is not changed in CPT-treated seedlings after any of the timepoints. Heat scale bar represents values of luminescence as counts per second.

Fig. 8 – ***asDOG1* expression changes in response to CPT in genAS plants**. Representative picture of 12-days old Col-0 seedlings carrying *pasDOG1-LUC::DOG1* (genAS) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *asDOG1* expression is not changed in CPT-treated seedlings after 24h (A) but is strongly increased in seedlings treated with 25 μM of CPT after 48h (B). Heat scale bar represents values of luminescence as counts per second.

Fig. 9 – ***asDOG1* expression changes in response to CPT in pAS plants**. Representative picture of 12-days old Col-0 seedlings carrying *pasDOG1::LUC* (pAS) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *asDOG1* expression

is not changed in CPT-treated seedlings after 24h (A) but is strongly decreased in seedlings treated with 10 and 25 μM of CPT after 48h. Heat scale bar represents values of luminescence as counts per second.

Fig. 10 – ***DOG1* expression changes in response to CPT in pSense plants**. Representative picture of 12-days old Col-0 seedlings carrying *psDOG1::LUC* (pSense) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *DOG1* expression is not changed in CPT-treated seedlings after 24h (A) but is slightly decreased in seedlings treated with 25 μM of CPT after 48h. Heat scale bar represents values of luminescence as counts per second.

Fig. 11 – **Framework of our DRIP-qPCR procedure** (adapted from Xu et al., 2017). Plant material are collected and used for nuclei isolation without any crosslinking step (opposite to standard ChIP), then genomic DNA is extracted and sonicated, and used for IP with the S9.6 monoclonal antibody. Next, hybrids are isolated using magnetic beads, and eluted for further qPCR analysis.

Fig. 12 – **R-loops detection within *DOG1 loci* in adult plants**. DRIP-qPCR on leaves of adult Col-0 plants with the selected primers for *DOG1* (*At5G45830.1*). Results shown as percent of input for samples not treated (Col-0) and treated with 7.5 U of recombinant *E. coli* RNase H (NEB, M0297S) overnight at 37°C (Col-0 + RNase H) as negative control. Strong signal is detected over the *DOG1* promoter region and exon2-intron2 junction. RNase H treatment prior IP decreased the signal. Bars show the average for three biological replicates, and error bars show the standard deviation. * show significant differences between the treated and not treated samples for each region of *DOG1*, with *t*-test for $p < 0.05$. On top is the schematic representation of DOG1 locus with the amplified region marked with green lines. Dashed lines match the amplified regions with the corresponding bars in the plot.

Fig. 13 – **R-loops detection within *DOG1 loci* in young seedlings**. DRIP-qPCR on 10-days old Col-0 seedlings with the selected primers for *DOG1* (*At5G45830.1*). Results show the percent of input for samples not treated (Col-0) and treated with 15 U of recombinant *E. coli* RNase H (NEB, M0297S) overnight at 37°C (Col-0 + RNase H) as negative control. RNase H treatment prior IP decreased the signal as expected. Strong signal is detected over the *DOG1* promoter region and exon2-intron2 junction. Bars show the average for three biological replicates, and error bars show the standard deviation. * show significant

differences between the treated and not treated samples for each region of *DOG1*, with *t*-test for $p < 0.05$. On top is the schematic representation of DOG1 locus with the amplified region marked with green lines. Dashed lines match the amplified regions with the corresponding bars in the plot.

Fig. 14 – **Model of R-loops assisting in the coordination of transcription bursts from sense and antisense *DOG1* promoters**. (A) Transcription bursts from *DOG1* sense promoter lead to transcription of *shDOG1* and *lgDOG1* transcript isoforms. Transcription of *lgDOG1* results in the readthrough of *asDOG1* promoter what is thought to mediate the repression of *asDOG1* transcription initiation. (B) After the complete round of sense transcription asDOG1 promoter is susceptible to be activated. Antisense transcripts possibly derived from *asDOG1* transcription form an R-loop over the sense promoter region which shut down *DOG1* transcription initiation and allows *DOG1* to bypass downstream *asDOG1*-mediated repression events. Once the R-loop is resolved by a specialized cellular machinery, sense transcription can be resumed. In the presence of CPT, R-loops formation is thought to increase, leading to a more frequent orchestration of sense and antisense transcription bursts. In this condition the optimal sense and antisense transcription is achieved. Transcription from the antisense promoter during sense transcription events would lead to conflicts such as PolII collisions, dsRNAs formation between the transcripts, competition for RNA binding proteins and other processing factors, etc.

Fig. S1 – **RNA samples considered to be of good quality**. Example of RNA samples ran on 1.2% agarose gel without signs of degradation and strong genomic DNA contamination. 100 ng of each RNA sample was loaded on the gel.

Fig. S2 – **DNA digestion confirmation**. Agarose gel image after PCR with primers for *PP2A* gene (*At1G69960.1*) on the RNA samples treated with DNase I. DNA ladder on the first lane, 6 samples run on lanes 2 to 7, and one positive control (RNA sample used on PCR not treated with DNase I) showing amplification of the genomic *PP2A* DNA sequence on the last lane, from left to right.

Fig. S3 – **Reanalysis of polyA site mapping by Direct RNA sequencing** (Sherstnev et al., 2012). Reads mapped to the antisense strand represent sites where polyadenylation occurs (*asDOG1* TTS) adapted from (Fedak et al., 2016).

Fig. S4 – **Strategy to identify the orientation of the RNA forming the R-loop at *DOG1* promoter**. Reverse transcription with the forward primer (Fw) leads to the synthesis of cDNA from an antisense transcript. Reverse transcription using the reverse primer (Rv)

leads to the synthesis of cDNA from a sense transcript. PCR with both primers on the Fw or Rv cDNA samples reveals the orientation of the RNA at the R-loop region.

Abbreviations and Acronyms

ABA - abscisic acid

asDOG1 - antisense *DELAY OF GERMINATION 1*

CHIP - Chromatin Immunoprecipitation

circRNAs - circularRNAs

CPT - camptothecin

DMSO - dimethyl sulfoxide

DNMT3B1 - DNA methyltransferase 3B1

***DOG1* -** *DELAY OF GERMINATION 1*

DRIP - DNA-RNA Immunoprecipitation

DRIP-qPCR - DNA-RNA Immunoprecipitation-quantitative Polymerase Chain Reaction

DRS - direct RNA sequencing

DSB - double-strand break

EM - Electron microscopy

ESCs - Embryonic stem cells

FLC - *FLOWERING LOCUS C*

H2Bubq - monoubiquitylation of histone H2B

H3K4me2 - dimethylation of histone 3 at lysine 4

H3K4me3 - trimethylation of histone 3 at lysine 4

H3K9Ac - acetylation of histone 3 at lysine 9

H3K27me1 - monomethylation of histone 3 at lysine 27

H3K27me3 - trimethylation of histone 3 at lysine 27

H3K36me3 - trimethylation of histone 3 at lysine 36

HUB1 - *HISTONE UBIQUITINATION 1*

IP - immunoprecipitation

lgDOG1 - long *DELAY OF GERMINATION 1*

lncRNA - long non-protein coding RNA

ncRNAs - non-coding RNAs

Pol I - DNA-dependent RNA polymerase I

Pol II - DNA-dependent RNA polymerase II

Pol IV - DNA-dependent RNA polymerase IV

qPCR - quantitative Polymerase Chain Reaction

PP2A - *SERINE/THREONINE PROTEIN PHOSPHATASE 2A*

PRC2 - Polycomb Repressive Complex 2

RdDM - RNA-directed DNA methylation

rDNA - ribosomalDNA

RT-qPCR - Reverse transcription-quantitative Polymerase Chain Reaction

shDOG1 - short *DELAY OF GERMINATION 1*

smFISH - single molecule Fluorescent *in situ* Hybridization

ssDNA - single-stranded DNA

ssDRIP-Seq - single-strand DNA ligation-based library construction from DNA-RNA immunoprecipitation followed by sequencing

TEs - transposable elements

TOP1 - *TOPOISOMERASE I*

TSS - Transcription start site

TTS - Transcription termination site

UBC21 - *UBIQUITIN-CONJUGATING ENZYME 21*

VIM - vimentin

Introduction

In eukaryotic organisms, the differential production of proteins that modulates growth and development is the result of well-orchestrated molecular mechanisms that regulate which genes are expressed in different cell types, stages of development or in response to different environmental stimuli. These mechanisms act for instance, in the remodeling of the chromatin structure, controlling transcription, RNA processing, RNA nuclear export and degradation, and at translational and post-translational levels.

"Waking up on time": Seed dormancy and germination

Plants are sessile organisms and their survival, and their reproductive success depend on the ability to perceive and respond to environmental signals. Seed dormancy is described as a mechanism that allows seeds to bypass temporarily unfavorable conditions. This allows seeds to align their germination with the environmental conditions that can support the entire plant's life cycle. Thus, the transition of dormant seeds to germination is the foremost decision seeds have to make. This is a very important and irreversible step, and so, seed dormancy and germination are extensively controlled processes that integrate endogenous and environmental signals (Finch-Savage and Leubner-Metzger, 2006). Abscisic acid (ABA) is a central endogenous player acting in the establishment and maintenance of seed dormancy (Finch-Savage and Leubner-Metzger, 2006). Mutations impairing ABA biosynthesis reduce seed dormancy (Leon-Kloosterziel *et al.*, 1996), whereas overexpression of biosynthesis genes enhance seed dormancy (Frey *et al.*, 1999; Xiong *et al.*, 2003). On the other hand, gibberellins display the antagonistic effect promoting germination. The dynamic balance between these hormones plays a central role in the regulation of seed dormancy/germination (Finch-Savage and Leubner-Metzger, 2006).

DELAY OF GERMINATION 1 (*DOG1*) was initially identified as a quantitative trait *locus* involved in the control of seed dormancy in *Arabidopsis thaliana*, a trait with high agronomical significance in many plants (Alonso-Blanco *et al.*, 2003). *DOG1* is the master regulatory gene of seed dormancy being absolutely required for the induction of dormancy in Arabidopsis seeds (Nakabayashi *et al.*, 2012). *DOG1* expression is tightly controlled and occurs in developing seeds, and decreases rapidly upon imbibition (Bentsink *et al.*, 2006).

DOG1 gene is regulated by its antisense transcript, named *asDOG1*. *asDOG1* is presumably a long non-protein coding RNA (lncRNA) transcribed from the second intron of *DOG1* in the antisense orientation (**Fig. 1**). This transcript is capped and polyadenylated, and negatively regulates *DOG1* expression. Moreover, *asDOG1* was shown to act in *cis* but not in *trans*, so that it has to be transcribed from the same DNA copy it acts on (Fedak *et al.*, 2016). Additionally, *DOG1* sense mRNA transcripts are alternatively spliced leading to the production of four mRNA isoforms (Bentsink *et al.*, 2006; Dolata *et al.*, 2015), and are also subject of alternative polyadenylation generating two transcript isoforms, the short isoform (*shDOG1*) and the long isoform (*lgDOG1*). Yet *shDOG1* is the one giving rise to a functional protein controlling the strength of seed dormancy (Cyrek *et al.*, 2016). *lgDOG1* transcription seems to function in the regulation of *shDOG1* transcription, and in the regulation of *asDOG1* through the monoubiquitylation of histone H2B (H2Bubq) (Kowalczyk *et al.*, 2017). *HISTONE UBIQUITINATION 1* (*HUB1*) encodes an E3 ligase enzyme that is essential for the deposition of the H2Bubq mark, and HUB1 was previously reported to regulate *DOG1* expression and seed dormancy levels (Liu *et al.*, 2007; Footitt *et al.*, 2015).

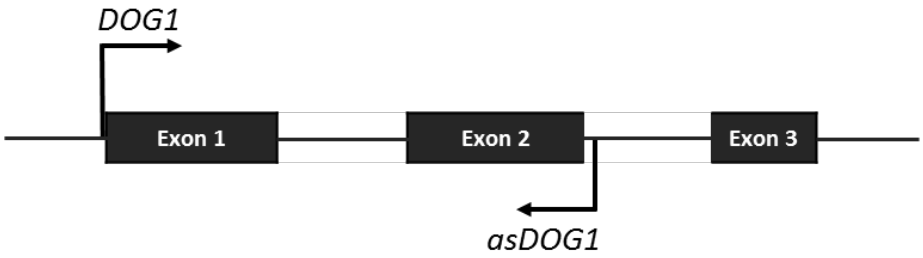


Fig. 1 – Schematic representation of *DOG1 locus*. Basic schematic representation of the *DOG1* (*At5G45830.1*) gene including its antisense transcript *asDOG1*. Exons (black boxes); introns in between, and upstream and downstream region (black lines); arrows show orientation of transcription (Fedak *et al*., 2016).

These findings reveal a tight control of *DOG1* long isoform transcription and H2Bubq mark deposition ultimately functions to orchestrate transcription of the functional *shDOG1* and *asDOG1* transcripts. *asDOG1* is thought to play an important role during seed maturation. At late stages of seed maturation *DOG1* expression decreases dramatically possibly due to the increase of *asDOG1* expression. In *asDOG1* mutants, *asDOG1* expression is compromised, and *DOG1* expression is higher throughout the seed development and is not downregulated as it occurs in wild type. This is followed by strong induction of DOG1 protein level and very strong seed dormancy phenotype (Cyrek *et al.*, 2016; Huo *et al.*, 2016). In addition to seed dormancy, *asDOG1* transcription was

shown to be crucial for the perception of drought/ABA signals in adult plants. In response to drought, *asDOG1* is downregulated by endogenous ABA signaling derepressing *DOG1* expression. Subsequently, *DOG1* transcripts levels increase, conferring tolerance to drought stress (Yatusevich *et al.*, 2017). Seed germination and drought tolerance are important agronomical traits, it is therefore imperative to understand the molecular regulation of *DOG1 locus*.

"A lovely attraction between RNA and DNA": R-loops as a novel class of regulators of gene expression

R-loops are DNA-RNA hybrids mainly formed co-transcriptionally by the annealing between a nascent RNA transcript and its complementary DNA template strand, leaving free a displaced single strand of DNA (**Fig. 2**; Roy and Lieber, 2009; Skourti-Stathaki and Proudfoot, 2014). The act of transcription itself generates negative supercoiling, which results in a more relaxed double helix state behind the transcribing RNA polymerase, offering the opportunity for the nascent RNA transcript to hybridize with its template strand, which results in the formation of a R-loop (Drolet *et al.*, 1995; Roy *et al.*, 2010; Aguilera and García-Muse, 2012). Besides negative supercoiling, RNA-DNA hybrids formation is also greatly influenced by the composition of the DNA sequence. Purine RNA-pyrimidine DNA duplexes were shown to be significantly more stable than DNA-DNA duplexes (Roberts and Crothers, 1992), and the superior thermodynamic stability of these hybrids is thought to drive R-loops formation *in vivo*.

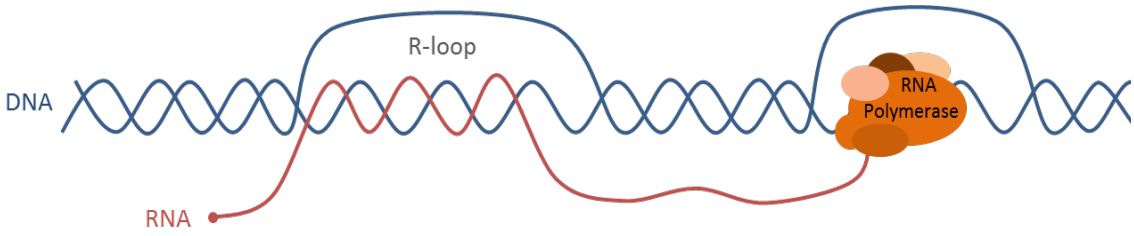


Fig. 2 – **Co-transcriptional R-loop formation**. R-loops formation when the nascent transcript invades the DNA double helix and hybridizes with its DNA template strand resulting in a structure composed by a DNA-RNA hybrid and a displaced strand of DNA that remains single stranded (structure so called R-loop; Aguilera and García-Muse, 2012).

In that sense Roy and Lieber (2009) revealed through *in vitro* experiments that the efficient formation of these structures relies on an initial step that requires G (guanosine) clusters and a following elongation that is primarily determined by G density, on the non-template DNA strand and C density on the complementary strand (referred as GC skew). More recently, the enrichment of R-loops was found to be associated not

only with GC skews but also with AT skews (Wahba *et al.*, 2016; Xu *et al.*, 2017) supporting that the thermodynamic stability of the DNA-RNA hybrids provided by their sequences influences R-loops formation *in vivo*. To counterbalance the over-accumulation of R-loops, cells possess specialized players that prevent and resolve these structures. The most well-known enzymes that resolve RNA-DNA hybrids once formed are RNase H which degrade the RNA hybridized to DNA (Drolet *et al.*, 1995; Aguilera and García-Muse, 2012). Other players include different helicases (Boule and Zakian, 2007; Mischo *et al.*, 2011; Skourti-Stathaki *et al.*, 2011; Sollier *et al.*, 2014; Song *et al.*, 2017; Cristini *et al.*, 2018), and DNA topoisomerases (Drolet *et al.*, 1995; El Hage *et al.*, 2010; Shafiq *et al.*, 2017), splicing factors (Li and Manley, 2005; Aguilera and García-Muse, 2012; Tanikawa *et al.*, 2016) and other mRNA processing and export factors (Huertas and Aguilera, 2003; Skourti-Stathaki and Proudfoot, 2014) that prevent R-loops formation, among other proteins (Bhatia *et al.*, 2014; Hatchi *et al.*, 2015; García-Rubio *et al.*, 2015).

R-loops formation have been for several decades linked with genomic instability derived from mutagenesis, recombination and chromosome rearrangement events (reviewed in Aguilera and García-Muse, 2012). However, R-loops have been recently found to contribute to the regulation of gene expression. The genome-wide mapping of R-loops in yeast, humans and plants revealed that these hybrids are highly abundant throughout the genome, and provide general insights on R-loops function in gene regulation, non-coding transcription and chromatin patterning, as well as predictive features of their formation (Ginno *et al.*, 2012; Sanz *et al.*, 2016; Wahba *et al.*, 2016; Xu *et al.*, 2017). In *Arabidopsis*, R-loops were found to be strongly enriched on promoter regions. Moreover, differences in these hybrids-forming regions between animals and plants were observed. In plants, the DNA-RNA hybrids were detected in low levels in the terminator regions, and detected at comparable levels in the sense and antisense orientations (Xu *et al.*, 2017) unlike what was observed in mammals (Sanz *et al.*, 2016). It is also important to notice that in plants R-loops were found enriched within the annotated lncRNAs regions, pointing towards a cross-talk between these two regulatory elements (Xu *et al.*, 2017). Despite the differences among different organisms, it is currently accepted that co-transcriptionally formed R-loops are prevalent structures in the genome with a strong sequence basis (GC skews and AT skews) and high transcription activity being a hallmark for their formation (Ginno *et al.*, 2012; Sanz *et al.*, 2016; Wahba *et al.*, 2016; Xu *et al.*, 2017). This raises the question of how R-loops once formed affect the next rounds of transcription.

"A portrait of R-loops on the transcriptional landscape": R-loops and transcription cross-talk

R-loops formation was firstly seen to impair the transcription elongation (Huertas and Aguilera, 2003; Tous and Aguilera, 2007). In fact, an extensive study on RNA polymerase I (Pol I) transcription in *S. cerevisiae* from Tollervey's laboratory strongly supports the function of R-loops in blocking transcription elongation. Upon Topoisomerase I (Top1) and Top2 depletion, the levels of RNA-DNA hybrids increased but this increase was much more dramatic in the quadruple mutant also lacking RNase H1 and H2 activities. Electron microscopy (EM) of Pol I transcription through ribosomal DNA (rDNA) genes revealed that Pol I piles up at the 5′ end of the 18S rDNA more frequently in strains lacking Top1 activity. Interestingly, the proportion of rDNA units with stalled Pol I was further increased in strains lacking both Top1 and RNase H activity. The electron microscopy observations revealed that Pol I piles up more strongly after depletion of Top1 and RNase H, correlating with the dramatic increased of R-loops accumulation in these strains, and support the idea that without the activity of these enzymes, the over-accumulation of R-loops together with the accumulation of positive supercoiling promotes Pol I to stall during transcription elongation along the rDNA genes (El Hage *et al.*, 2010).

In addition to transcription elongation and R-loops feedback, these hybrids formation was also found to interplay with mRNA splicing. The first evidences supporting the regulatory role of splicing over the formation of R-loops came from studies in which the depletion of different processing and export factors induced R-loops formation mediating increased genomic instability (Huertas and Aguilera, 2003; Li and Manley, 2005). A recent study in *Saccharomyces cerevisiae* showed that intron-containing genes display decreased levels of R-loops and DNA damage, in contrast to intron-less genes. Moreover, the insertion of introns in R-loop-prone genes attenuates R-loops accumulation and, subsequently, transcription-associated genomic instability. Interestingly, this protective effect of introns was shown to be due to the recruitment of the spliceosome machinery itself and not the result of splicing; it is therefore likely that the recruitment of splicing factors to the DNA hampers the hybridization of the nascent RNA transcript to its DNA template (Bonnet *et al.*, 2017). This preventive function of the spliceosome is consistent with previous findings that many mRNA processing and export factors prevent the formation of R-loops (Li and Manley, 2005; Aguilera and García-Muse, 2012; Tanikawa *et al.*, 2016). In addition to the exhaustive regulation of R-loops

formation by RNA splicing and processing, alternative splicing was also found to be regulated by R-loops in Arabidopsis. The mechanism however is rather unusual, in which circular RNAs (circRNAs) derived from *SEPALLATA3*'s (*SEP3*) exon 6 bind to their complementary DNA sequence forming R-loops in *trans*, thus promoting the biogenesis of the exon 6-skipped mRNA isoform (Conn *et al.*, 2017).

Contrasting to the first notion that R-loops impair transcription, other studies revealed that the formation of these hybrids play a critical role for the proper termination of transcription in human cell lines (Skourti-Stathaki *et al.*, 2011, 2014). In fact, the transient formation of R-loops in transcription pause sites downstream of poly(A) sites seems to induce the pausing of RNA Polymerase II (Pol II). Then, these R-loops must be resolved by the RNA/DNA helicase Senataxin to allow the degradation of the nascent RNA by Xrn2 5′–3′ exonuclease, that ultimately promotes the efficient termination of transcription (Skourti-Stathaki *et al.*, 2011). Few years later, a different mechanism of transcription termination was proposed. Strong evidences support the idea that the formation of these hybrids over the pause elements prior transcription termination lead to antisense transcription from the displaced ssDNA, and subsequently generation of double-stranded RNAs (dsRNAs). These dsRNAs then recruit the RNAi machinery, which include Dicer, Ago1, Ago2, and G9a methyltransferase, which in turns lead to the deposition of H3K9me2 repressive marks over those regions. H3K9me2 and Heterochromatin Protein 1γ (HP1γ) recruitment then reinforce Pol II pausing prior to efficient transcription termination. Interestingly, this mechanism of R-loops-mediated transcription termination involving the recruitment of the RNAi apparatus seems to be widespread in the human genome (Skourti-Stathaki *et al.*, 2014). Whether in *Arabidopsis*, R-loops play any role in transcription termination is still an open question, though the enrichment of these hybrids in terminator regions is much lower comparing with the enrichment in mammalian terminators (Sanz *et al.*, 2016; Xu *et al.*, 2017). Nevertheless, this mechanism opens a new perspective on what could be the relationship between R-loops and antisense transcription.

"A matter of orientation": Antisense transcription and R-loops formation

The recent mapping of R-loops spatial distribution throughout the Arabidopsis genome revealed several new insights on the connection between R-loops and antisense transcription. For instance, 39.6% of genes were found to have both sense

and antisense R-loops and 21.5% of genes were found to have antisense R-loops only (Xu *et al.*, 2017). These numbers clearly point towards a relationship between these two regulators.

As discussed previously, R-loops in human cell lines were found to be transiently formed over transcription termination regions associated with the synthesis of small antisense transcripts generated from the displaced strands of DNA in the R-loops (Skourti-Stathaki *et al.*, 2014). This exposes a mechanism in which R-loops can trigger antisense transcription without the presence of a promoter. It is interesting to think that the displacement of a single strand of DNA on the R-loop offers the opportunity of that strand to be transcribed and generate antisense transcripts. We can speculate that these R-loops-derived antisense transcription events may also play other functions in addition to transcription termination.

Not only R-loops mediate antisense transcription, but antisense transcripts can also induce R-loops formation. Several studies have been revealing the role of transcription of non-coding RNAs (ncRNAs) in double-strand break (DSB) repair in eukaryotes (Wei *et al.*, 2012; Gao *et al.*, 2014; Keskin *et al.*, 2014). In *Schizosaccharomyces pombe*, a recent study addressed the role of R-loops in this process (Ohle *et al.*, 2016). The authors found that Pol II-derived transcripts form DNA-RNA hybrids around the DSB sites. Interestingly, it was shown that RNase H activity is necessary for efficient homologous recombination, and both deletion and overexpression of this enzyme impair the process of repair (Ohle *et al.*, 2016). The results suggest that the R-loops formation is necessary, yet it is also required their degradation by RNases H to complete the DSB repair. From both DSB repair (Ohle *et al.*, 2016) and transcription termination (Skourti-Stathaki *et al.*, 2011, 2014) studies it is worth noting that the transient formation of R-loops is an important feature mediating their function.

In *Arabidopsis thaliana*, the formation of an R-loop in the promotor region of the antisense lncRNA *COOLAIR* at the *FLC locus* allows one ssDNA to remain free and susceptible to the binding of NDX, a single-stranded DNA (ssDNA) binding protein. The binding of NDX therefore stabilizes the R-loop structure decreasing the levels of transcription of *COOLAIR* what mediates the increase of sense *FLC* expression (Sun *et al.*, 2013). A different mechanism was reported in humans, where the transcription of the antisense transcript from *VIM locus* leads to the formation of an R-loop in the promoter region of *VIM* gene which is associated with a decrease in the nucleosome occupancy and enhanced binding of transcription factors that are positive regulators of *VIM*

expression (Boque-Sastre *et al.*, 2015). In agreement, a previous study found the formation of R-loops associated with chromatin decondensation at the *Snord116 locus*. The R-loops formation was found to repress transcription of *Ube3a* antisense transcript, reverting the paternally imprinted silencing of the *Ube3a* sense expression that drives an autism-spectrum disorder (Powell *et al.*, 2013). These later examples not only highlight the strong relationship between R-loops and antisense transcription but also suggest that R-loops may have relevant functions in modifying chromatin architecture.

"R-loops shaping the chromatin silhouette": The link between R-loops and chromatin remodeling

The mapping of R-loops spatial distribution throughout the Arabidopsis genome showed the colocalization of these hybrids with regions with active histone marks such as H3K36me3, H3K4me2, H3K4me3, H3K9Ac and H3K27me1 (Xu *et al.*, 2017). This was expected since R-loops are mainly formed during transcription, and their formation correlates with high gene expression (Ginno *et al.*, 2012; Sanz *et al.*, 2016; Wahba *et al.*, 2016; Xu *et al.*, 2017). Nonetheless, the previously described examples at the *VIM* (Boque-Sastre *et al.*, 2015) and *Snord116 loci* (Powell *et al.*, 2013) suggest that augmented chromatin decondensation can also be the consequence of R-loops formation rather than the opposite.

In Arabidopsis, it was surprising though, that R-loops were also found strongly enriched in regions with the heterochromatin histone mark H3K9me2 (Xu *et al.*, 2017). Indeed, using a DNase I sensitivity approach to assess chromatin accessibility, H3K9me2 marked regions were found to be the least accessible chromatin regions in Arabidopsis (Shu *et al.*, 2012). So, what explains the colocalization of R-loops-forming regions with heterochromatin regions marked with H3K9me2? As discussed by Qianwen Sun and his colleagues (2017), one possibility is that R-loops may participate in a mechanism that leads to the deposition of this repressive histone mark. Since they also found a strong enrichment of R-loops formation within intergenic RNA polymerase IV (Pol IV)-transcribed noncoding regions and transposable elements (TEs) it was suggested that R-loops may function in the plant specific pathway of RNA-directed DNA methylation (RdDM; Xu et al., 2017). This putative function in mediating transposons silencing through RdDM can be supported in light of the knowledge about R-loops function mediating transcription termination by the deposition of H3K9me2 in mammals (Skourti-Stathaki *et al.*, 2014).

In mouse embryonic stem cells (ESCs), R-loops were found enriched at promoter regions bound by the Tip60-p400 histone acetyltransferase complex, and overexpression of RNase H reduced the accumulation of RNA-DNA hybrids and decreased the binding of Tip60 and p400 to most Tip60-p400-target genes. Moreover, R-loops were poorly enriched near the promoter proximal regions of genes highly bound by Suz12, a subunit of the Polycomb Repressive Complex 2 (PRC2), and RNase H overexpression increased Suz12 binding to its target genes and off-target genes, and increased methylation of histone 3 at lysine 27 (H3K27me3). Together, this study revealed that the formation of promoter-proximal R-loops promotes the binding of Tip60-p400 complex and prevents the binding of PRC2 and subsequent H3K27me3 deposition to those chromatin regions. The regulation of Tip60-p400 and PRC2 binding at promoter regions by R-loops formation was suggested to be a general role by which R-loops enable ESCs to efficiently respond to differentiation cues (Chen *et al.*, 2015).

Despite of the extensive interplay with the chromatin decoration with specific histone modifications, this is not the only link between R-loops and the *epi* layer of regulation of gene expression. R-loops formation was also found to be involved in remodeling chromatin environment through DNA methylation. In human CpG islands, R-loops formed upon transcription is suggested to protect these promoters from DNMT3B1-mediated DNA methylation (Ginno *et al.*, 2012). In agreement, R-loops formation was found to negatively correlate with CG DNA hypermethylation in Arabidopsis (Xu *et al.*, 2017), suggesting that the promoter R-loops that function as a shield against DNA methylation in humans may be conserved in plants. These observations are consistent with genome-wide increased chromatin accessibility associated with R-loops formed around the transcription start site (TSS), as reported in humans (Sanz *et al.*, 2016).

The recent efforts to study the R-loops formation and its functions have illustrated the large spectrum of actions of these regulators. Currently, we perceive R-loops not only as byproducts of transcription associated with genomic instability but also as heavyweight players in the control of gene expression cooperating with several regulatory elements and integrating various layers of regulation.

In this study, DNA-RNA Immunoprecipitation (DRIP) assays were used to map R-loops formation within the *DOG1 locus*, and differential expression studies were conducted to address their function in the regulation of sense and antisense expression at this *locus*.

Methods

Plant material and growth conditions

Arabidopsis thaliana wild-type Columbia (Col-0; Swiezewski's laboratory collection) plants were used as plant material for the experiments performed in this work. Additionally, for the luciferase reporter assay were used the transgenic lines *pasDOG1-LUC::DOG1* (unpublished), *pasDOG1::LUC*, *psDOG1::LUC::DOG1* and *psDOG1::LUC* (Fedak *et al.*, 2016) on Col-0 background generated before by other members of the laboratory. Plant material was collected always around 10 am, and immediately frozen in liquid nitrogen and stored at -80 °C.

Seeds sterilization

Arabidopsis seeds were surface-sterilized by vapor-phase sterilization following the steps described hereafter. Opened Eppendorf tubes with around 100 µL of seeds were placed inside a desiccator jar together with a flask with 100 mL of sodium hypochlorite. Then, 10 mL of HCl was added carefully to the flask with sodium hypochlorite and the desiccator jar was sealed to allow the seeds to be sterilized by the chlorine gas for 2 h.

In vitro culture

Surface-sterilized seeds were sown in Petri dishes with ½ MS solid medium (½ Murashige-Skoog (Sigma Aldrich), 0.7% (w/v) plant agar (Duchefa), pH 5.7-5.8). Seeds were stratified at 4°C for two days, and then transferred to a growth chamber with long day conditions (16 h light/8 h darkness) at 22°C /18°C. CPT (Sigma-Aldrich; C9911) was added to autoclaved media just before preparing the plates.

Ex vitro culture

Pots 13x13x13 cm (Interplast Plastic Products BYTOM) were previously prepared with soil and watered. Surface-sterilized seeds were sowed on the soil, and the pots were placed at 4°C for two days covered to keep the moisture. Seven to ten days later, individual seedlings were transferred to new pots. Plants were grown for 6 weeks in a glasshouse under controlled environmental parameters: 70% humidity, temperature 22°C, 16 h light/8 h dark photoperiod regime at 150–200 mE/m^2.

Luciferase reporter assay

Seedlings were sprayed with a solution with 0.5 mM beetle luciferin (Promega) and 0.015% Silwet L-77 (Lehle Seeds) and placed under a NightShade LB985 camera. Photos for data visualization were processed using IndiGO (Berthold) imaging software. *pasDOG1-LUC::DOG1* (unpublished), *pasDOG1::LUC*, *psDOG1::LUC::DOG1* and *psDOG1::LUC* (Fedak *et al.*, 2016) were used. Experiments were performed with at least 3 replicates for each treatment, with at least 12 seedlings per replicate.

RNA extraction

Plant material was ground into a fine powder in liquid nitrogen using a pestle and mortar. Total RNA was extracted by hot phenol method described in (Shirzadegan *et al.*, 1991). Five hundred µL of homogenization buffer (100 mM Tris pH 8.5; 5 mM EDTA pH 8; 100 mM NaCl; 0.5% (w/v) SDS) with 5 µL of β-mercaptoethanol (Sigma Aldrich) were added to each sample. Then, 250 µL of phenol (Applichem) at 60°C was added to each sample and tubes were shaken for 15 min at 60°C at 1400 rpm on an Eppendorf Thermomixer Comfort (Eppendorf AG, Hamburg, Germany). Then, 250 µL of chloroform (POCH) was added to each sample, and samples were shaken for 15 min at room temperature at 1400 rpm, and centrifuged for 10 min at 20,817 *g*. The top aqueous layer from each sample was transferred to new tubes and 550 µL of a phenol:chlorofolm:isoamyl alcohol (25:24:1) mixture (Sigma Aldrich) was added. Tubes were incubated for 10 min at room temperature at 1400 rpm, and then centrifuged for 10 min at 20,817 *g*. Five hundred µL of the top aqueous layer was carefully transferred to fresh Eppendorf tubes, and 50 µL of 3 M sodium acetate, and 400 µL isopropanol were added. The tubes were incubated for 30 min at -80 °C, and then centrifuged for 30 min at 4°C at 20,817 *g*. The pellet was washed in 400 µL of 80% ethanol, centrifuged for 10 min at 4 °C at 20,817 *g*, and the ethanol was removed. The pellet was air-dried for 7 min, dissolved in 40 µL of sterile miliQ H_2O, and left overnight at 4°C. Quality of RNA was firstly examined by electrophoresis in 1.2% (w/v) agarose gel. Purified RNA was stored at -20 °C for short term storage.

In order to remove DNA contamination, RNA samples were treated with TURBO DNA-*free* ™ kit (Thermo Fisher Scientific) according to the manufacturer's instructions for rigorous DNase treatment procedure with the only modification that the samples were incubated with TURBO DNase for 30 min initially, plus 20 min in the second incubation.

Lack of DNA in the RNA samples was confirmed by PCR. RNA concentration and quality were measured using a Nanodrop 2000 spectrophotometer.

Nuclei isolation

For nuclei isolation 2 g of plant material was ground into a fine powder in liquid nitrogen using a pestle and mortar. The powder was mixed with 15 mL of Honda Buffer (20 mM HEPES-KOH pH 7.4, 0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT) on 50 mL falcon tubes for 10 min at 4°C and then filtered through two layers of Miracloth and centrifuged at 2000 g for 15 min. Nuclear pellets were washed once with 1 ml of Honda buffer.

DNA extraction

DNA extraction from lysed nuclei was performed by adding 300 µL of phenol pH 7.8-8.2 (Applichem) to 300 µL of lysed nuclei followed by 15 sec vortexing and a 5 min centrifugation at 20,817 *g*. Then, 300 µL of the supernatant were transferred to new tubes and 300 µL of phenol pH 7.8-8.2 (Applichem) were added for the second time followed by 15 sec vortexing and a 5 min centrifugation at 20,817 *g*. Then, 300 µL of the supernatant were transferred to new tubes and 300 µL of chloroform (POCH) were added, followed by 15 sec vortexing and a 5 min centrifugation at 20,817 *g*. The supernatant was transferred to new tubes and incubated with 30 µL of 3 M sodium acetate and 900 µL of ethanol 96% for at least 1 h at -80°C. The samples were centrifuged for 30 min at 4°C at 20,817 *g*. The pellet was washed with 900 µL of 70% ethanol, centrifuged for 10 min at 4°C at 20,817 *g*, and the ethanol was removed. The pellet was air dried for 10 min and dissolved in 80 µL of sterile miliQ H_2O.

cDNA synthesis

cDNA synthesis was performed by reverse transcription with heat denaturation according the Two-Step RT-PCR Procedure using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) according to manufacturer's instructions. 2.5 µg of RNA was used for reverse transcription. First the RNA template was mixed with the RT primers (1 µL of 100 µM Oligo$(dT)_{18}$ or 15 pmol of gene-specific primer) and sterile miliQ H_2O up to 12 µL. The samples were incubated at 65°C for 5 min and placed immediately on ice. Then, to each sample 4 µL of 5X Reaction Buffer, 1 µL of RiboLock RNase Inhibitor (20 U/µL), 2 µL of 10 mM dNTP Mix and 1 µL of RevertAid M-MuLV RT (200 U/µL) were added for a final volume of 20 µL. For Oligo$(dT)_{18}$-primed cDNA

synthesis the reaction was performed for 60 min at 42°C followed by 15 min at 95°C. For gene-specific-primed cDNA synthesis the reaction was performed for 30 min at 50°C followed by 15 min at 95°C. The cDNA was diluted by adding 30 µL of sterile miliQ H_2O, and used in PCR or stored at -20°C.

Primers design

Primers were designed using Primer 3 Plus (https://primer3plus.com/) and Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) taking in account the following parameters: primer size around 20 bp; T_m around 60°C, similar between each forward and reverse primer; % GC around 50; amplicon size between 60 and 200; and lowest self and pair complimentary and secondary structures possible.

Polymerase chain reaction (PCR)

PCR amplification was performed in 20 µL total volume on 0.2 mL tubes with a mixture of 1 µL of each 10 µM primer, 10 µL of 2X DreamTaq Green PCR Master Mix (Thermo Fisher Scientific), 1 µL of template, and 7 µL of sterile miliQ H_2O. The applied program consisted in 4 min for initial denaturation at 94°C followed by 30 cycles of 15 s at 94°C (denaturation), 15 s at 55°C (annealing), 15 s at 72°C (extension) and final extension for 2.5 min at 72°C. The list of all primers used in this study is given in Table S1.

Quantitative PCR (qPCR)

Quantitative PCR was performed using a LightCycler 480 Instrument (Roche). Reactions were performed in 20 µL final volume with 0.5 µL of each 10 µM primer, 10 µL of SYBR Green Master Mix (Roche), template, and sterile miliQ H_2O up to 20 µL, on LightCycler 480 384 Multiwell Plate (Roche) covered with LightCycler 480 Multiwell Sealing Foil (Roche). Amplification program included an initial activation step at 95°C for 10 min, followed by 45 cycles of denaturation at 95°C for 10 s, primers annealing at 58°C for 15 s and extension at 72°C for 15 s, finally an extension step at 72°C for 10 min and a final dissociation curve step at 95°C for 8.5 min. The calculations were performed using Microsoft® Excel 2013. The results were normalized against the expression of reference gene *UBC* (*At5g25760*). The list of all primers used in this study is given in Table S1. All qPCR experiments were done with 3 biological and 3 technical replicates.

Agarose gel electrophoresis

Electrophoretic separation of nucleic acids was performed on agarose gels with 10 µL of ethidium bromide (10 mg/mL) per 50 mL gel in TBE buffer (0.04 M Tris-Borate; 0.1 mM EDTA pH 8). Electric potential of 75 V was applied in the electrophoresis for the appropriate amount of time (depending on the size of the fragments under analysis and on the amount of separation required). The size of the DNA bands was estimated by loading in the same gel 7 µL of GeneRuler DNA Ladder Mix (SM0331) (Thermo Fisher Scientific). The images of the gels were captured with GelDoc G:BOX EF2 (Syngene) and analyzed using the GeneSys image acquisition software (Syngene).

DNA-RNA Immunoprecipitation (DRIP)

The isolated nuclei from 2 g of plant material was resuspended in 300 µL of Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% (w/v) SDS), followed by 1 h incubation with 3 µL RNase A (Thermo Fisher Scientific) at 37°C, and for 6 h with 8 µL proteinase K (Thermo Fisher Scientific) at 50°C, and following DNA extraction. DNA was sonicated with Covaris™ (twice 40 sec with settings at duty cycle: 20%, intensity: 10, cycles/burst: 200, and 30s rest in between) in microTube AFA Fiber 6x16mm (Covaris) to get fragments between 400 and 500 bp. The sonicated DNA was treated with 7.5 or 15 U, depending on the experiment, of recombinant *E. coli* RNase H (NEB, M0297S) in a final volume of 40 µL, and incubated overnight at 37°C. From all samples 10% of the volume was taken just before IP, diluted in the same final volume as the eluted samples after IP and used for qPCR. Both samples: treated and not treated with RNase H, were mixed with 360 µL of IP Buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl) and 2 µg of Anti-DNA-RNA Hybrid [S9.6] antibody (Kerafast) in 1.5 mL maximum recovery tubes (Corning Axygen), and incubated overnight at 4°C. Then 8 µL of Dynabeads Protein G (Thermo Fisher Scientific) were mixed with 400 µL of IP Buffer in maximum recovery tubes, and vortexed. Then the tubes were placed in a magnetic rack for 1 min, and the buffer was removed. The same amount of IP Buffer was added again, and the washing was repeated once more. After the washing steps the samples incubated with the antibody were transferred to the tube with the beads. The samples were briefly vortexed to mix the beads and incubated for 4 h at room temperature on a rotation wheel. Bead-antibody complexes were washed 5 times 5 min each. Twice with Low Salt Buffer (0.5% Triton X-100, 20 mM Tris-HCl pH 8, 0.2% (w/v) SDS, 2 mM EDTA, 150 mM NaCl), once with High Salt Buffer (0.5% Triton X-100, 20 mM Tris-HCl pH 8,

0.2% (w/v) SDS, 2 mM EDTA, 500 mM NaCl), and twice with TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). The washing steps consisted in placing the samples in the magnetic rack, waiting for 1 min, removing the buffer and adding the new washing buffer immediately after, and mixing the beads by brief vortexing followed by an incubation for 5 min on a rotation wheel. After washing 100, µL of 10% Chelex-100 (Bio-Rad) were added to the bead-antibody complexes, and incubated 10 min at 95°C. Then, the samples were placed at room temperature for 5 min and 2 µL of proteinase K (Thermo Fisher Scientific) were added followed by 30 min incubation at 50°C. Then, samples were incubated again for 10 min at 95°C, centrifuged for 5 min at 20,817 *g* at room temperature. Seventy µL of the supernatant containing the eluted DNA-RNA hybrids were transferred to new tubes, and 50 µL of miliQ H_2O were added to the remaining of the centrifuged samples to resuspend the pellets. These were centrifuged again, and 50 µL of the supernatant was added to the previously transferred 70 µL. The eluted samples treated and not treated with RNase H in a final volume of 120 µL together with the 10% input samples and “no antibody” samples were used for qPCR. The “no antibody samples” were obtained using miliQ H_2O instead of the S9.6 antibody.

Aims

As mentioned before, the *DOG1* gene displays a keystone role in seed dormancy/germination and in drought stress tolerance in adult Arabidopsis plants. As such an important gene, *DOG1* is extensively regulated at different layers. This work focused on getting new insights on a putative novel regulatory mechanism at the transcriptional level involving R-loops as central players. In the pursuit of R-loops function in the regulation of the *DOG1 locus* the following aims were highlighted: (i) to understand how *DOG1* and *asDOG1* expression are changed in a condition where R-loops formation is changed. This will provide the first insights on how R-loops regulate sense and antisense expression at the *DOG1 locus*; (ii) to explore the interplay between *asDOG1* transcription and the R-loops functions in gene expression at the *DOG1 locus*. This will provide a broader picture of R-loops functions and regulation; and (iii) to establish a technique to allow the mapping of R-loops within the *DOG1 locus*. This will enable detection of R-loops formation throughout the *DOG1 locus* what will help to understand the possible roles R-loops may play in the regulation of *DOG1 locus*.

Results and Discussion

CPT-mediated changes in *DOG1* sense and antisense expression levels

Recently, Qianwen Sun and his colleagues developed a novel variant of DRIP for the genome-wide mapping of R-loops, and described for the first time the detection of R-loops formation throughout the *Arabidopsis thaliana* genome (publication from Qianwen Sun group, Xu et al., 2017). The sequencing data showed enrichment of mapped reads over two regions of the *DOG1 locus*, on the promoter region and on *DOG1* first intron (**Fig. 3**). Moreover, these two R-loop-forming regions are formed one on the sense and the other on the antisense strand (Xu *et al.*, 2017). This, together with the fact that R-loops have been found to play various roles in the regulation of gene expression prompted us to investigate their function in the regulation of sense and antisense expression within the *DOG1 locus*.

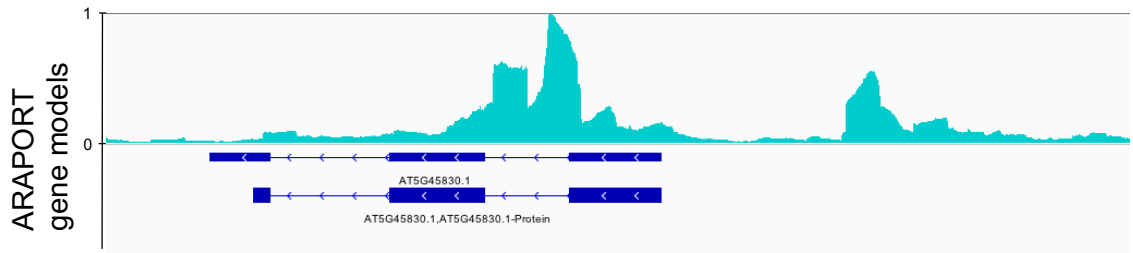


Fig. 3 – **R-loop formation within *DOG1 locus* detected by ssDRIP-Seq**. Snapshot of the ssDRIP-seq data on *DOG1* genomic region (*At5G45830.1*) from the Arabidopsis genome-wide mapping of R-loops described in (Xu et al., 2017). Strong R-loops detection over *DOG1* promoter region and *DOG1* intron 1. y axis represents R-loops normalized reads number in auxiliary units. Gene annotation in the bottom (Xu et al., 2017).

For that, the first theoretical approach was to determine the differential sense and antisense *DOG1* transcripts' levels between plants displaying the normal R-loops formation and plants displaying increased R-loops formation. There is no characterized Arabidopsis mutant displaying either increased or decreased genome-wide R-loops formation. Nonetheless, camptothecin (CPT) can be used for this purpose. CPT is a plant alkaloid that does not only bind to TOP1, but also stacks between the base pairs that flank the TOP1-associated DNA cleavage site sequestering TOP1 cleavage complex in a reversible state (Pommier, 2006). CPT-inhibition of TOP1 ability to release the supercoiling during transcription was used before in several manuscripts as a strategy to increase the formation of co-transcriptionally formed R-loops, including plants (El

Hage *et al.*, 2010; Powell *et al.*, 2013; Groh *et al.*, 2014; Marinello *et al.*, 2016; Shafiq *et al.*, 2017). Based on this, the first experimental approach was to determine the effect of CPT on the steady-state *DOG1* and *asDOG1* transcripts levels through RT-qPCR on wild-type *Arabidopsis thaliana* Col-0 seedlings not treated versus treated with CPT.

Col-0 seedlings were grown for 5 days on solid half-strength MS media and transferred to media with 2 µM of CPT, 25 µM of CPT or DMSO as control. The plant material was used for RNA extraction with phenol-chloroform, and the quality of the RNA was first verified by agarose gel electrophoresis. The quality was considered good to proceed if the rRNA bands were sharp without smear and without considerable genomic DNA or protein contamination (**Fig. S1**). The extracted RNA was first subjected to DNase I treatment to degrade the genomic DNA that could bias the qPCR results. The efficient DNA digestion was confirmed by performing a PCR on the RNA samples after the DNase I treatment with primers for a genomic fragment of *PP2A* gene (*At1G69960.1*) and analyzed on a gel (**Fig. S2**). Lack of visible bands indicated successful removal of genomic DNA. After the DNA digest the RNA samples were ran once again on a gel to confirm their integrity, and the concentration was measured by NanoDrop (Thermo Scientific™). Since the *asDOG1* RNA sequence is complementary throughout its full length to the *DOG1* RNA transcripts and genomic DNA sequences, for detection of *asDOG1, asDOG1*-specific primers with adapters were used in reverse transcription step allowing to specifically detect in qPCR *asDOG1* RNA despite the presence of the complementary *DOG1* sense mRNA (**Fig. 4**; Fedak et al., 2016).

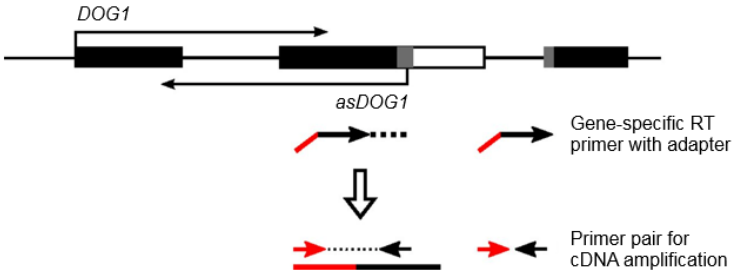


Fig. 4 – ***asDOG1* reverse transcription strategy**. Schematic diagram of the *DOG1* (*At5G45830.1*) gene showing the gene-specific first strand cDNA synthesis using a primer with an *asDOG1*-complementary sequence tagged with an adapter sequence at its 5' end, and specific PCR amplification of *asDOG1* cDNA using the adapter sequence as forward primer. Exons are represented by black boxes; introns in between, and upstream and downstream region (black lines); exon 2 in the *shDOG1* is extended (white box); exonic regions derived from alternative splicing (grey boxes) (adapted from Fedak et al., 2016).

RT-qPCR results showed a not statistically significant two-fold increase of *DOG1*, and a significant three-fold increase of *asDOG1* in seedlings grown in the presence of 25 µM of CPT for 36h (**Fig. 5**). *AtGP1* was used as a positive control since it was published to be overexpressed upon an identical experimental exposure to CPT as performed in this study (Dinh *et al.*, 2014). The increases detected for both *DOG1* and *asDOG1* in the presence of CPT came as a surprise since in other studies conducted in the same laboratory where this analysis was performed the increase of *asDOG1* expression was routinely associated with a decrease of *DOG1* expression and vice versa.

Next, a series of measurements of the luminescence levels in transgenic lines containing the Luciferase reporter fused to *DOG1* or *asDOG1* in the genomic context or to separated promoters as shown in **Figure 6** were performed. For the reporter assay, 12-days old seedlings growing in MS media were transferred to MS plates containing DMSO (control), 2 µM, 10 µM and 25 µM of CPT, and the intensity of luminescence was measured 24 and 48 hours after the transfer.

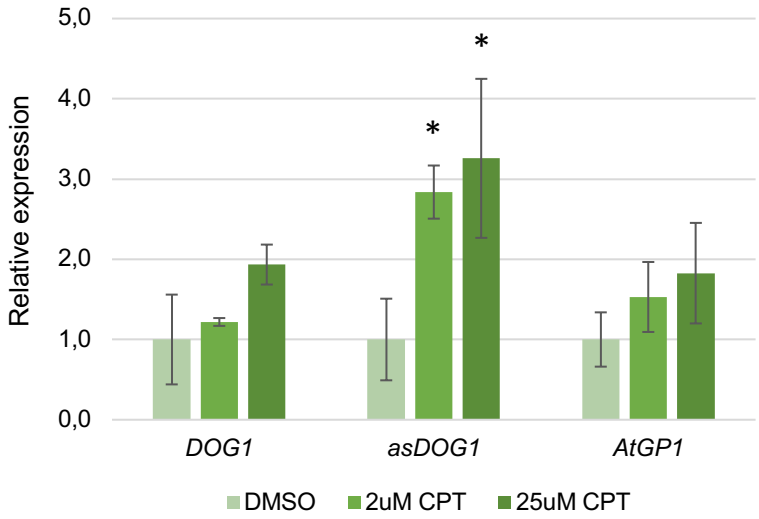


Fig. 5 – **CPT effect on sense and antisense *DOG1* expression**. RT-qPCR for *DOG1*, *asDOG1* (*At5G45830.1*) and *GP1* (positive control; Dinh et al., 2014) in Col-0 seedlings growing in the presence of 2 µM of CPT 25 µM of CPT or DMSO as control. Expression levels were normalized against *UBC21* (*At5G25760.1*) mRNA; data represents the means of three biological replicates for each treatment condition with error bars representing standard deviation. * show significant differences with *t*-test for $p < 0.05$.

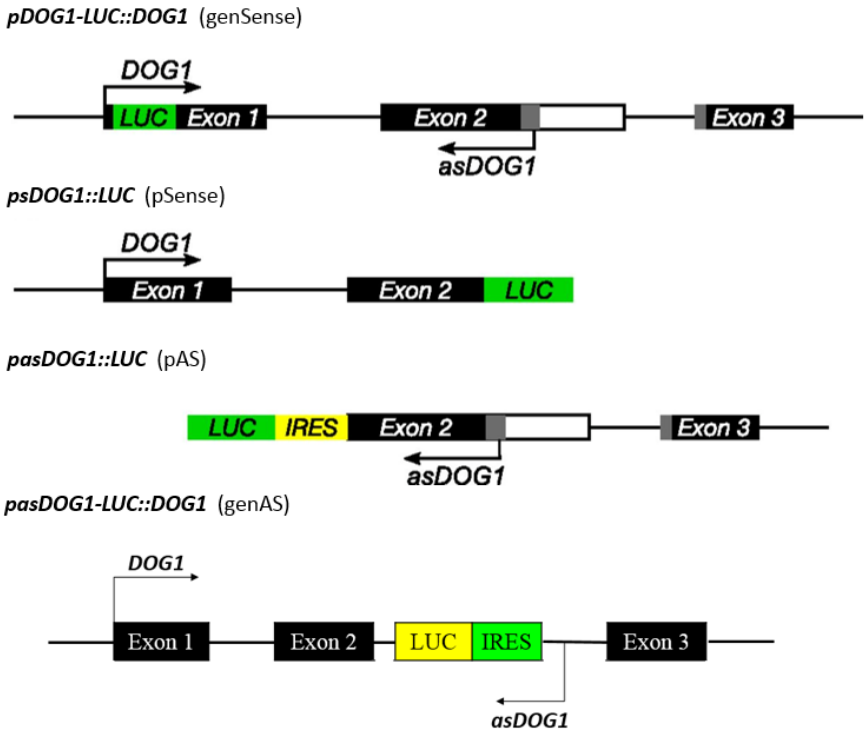


Fig. 6 - Schematic diagrams of the reporter constructs – Constructs used to obtain the reporter transgenic lines (performed before by members of the laboratory) containing the luciferase reporter cassette fused to *DOG1* in the genomic context (*pDOG1-LUC::DOG1* referred as genSense), fused to *DOG1* in separated promoter (*pDOG1::LUC* referred as pSense), fused to *asDOG1* in separated promoter (*pasDOG1::LUC* referred as pAS) (Fedak et al., 2016), and fused to *asDOG1* in the genomic context (*pasDOG1-LUC::DOG1* referred as genAS) (unpublished). The *asDOG1* constructs contain an additional IRES sequence to drive translation of the RNA transcripts.

The luciferase reporter assay performed on *psDOG1-LUC::DOG1* (hereafter referred as genSense) seedlings subjected to CPT treatment showed no changes in the luminescence levels between CPT-treated seedlings and control at any timepoint (**Fig. 7**). These results contrast with the RT-qPCR results (**Fig. 5**) showing *DOG1* overexpression in response to CPT. While RT-qPCR analysis allow the quantification of RNA transcripts, the Luciferase reporter assay allow the gene expression quantification at the protein level. However, it is unlikely that this would be the reason for the different results obtained with the different methods since TOP1 is the only known cellular target of CPT, and no effect is expected at the translational level (Pommier, 2006). Nonetheless it is possible that the insertion of the Luciferase reporter gene in the construct may have made the transgene unable to respond to CPT.

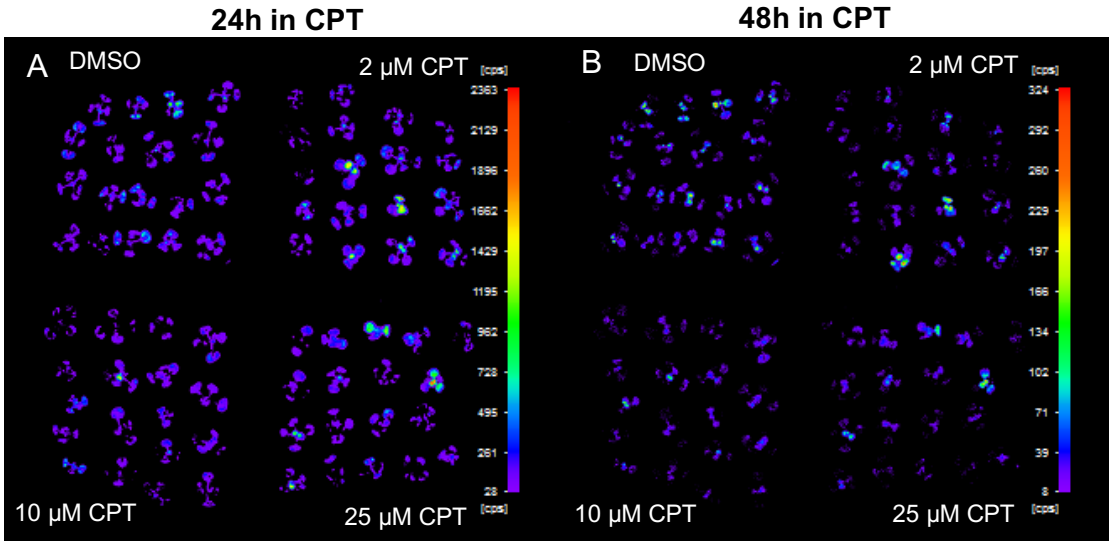


Fig. 7 – ***DOG1* expression changes in response to CPT in genSense plants**. Representative picture of 12-days old Col-0 seedlings carrying *psDOG1-LUC::DOG1* (genSense) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *DOG1* expression is not changed in CPT-treated seedlings after any of the timepoints. Heat scale bar represents values of luminescence as counts per second.

The measurements performed using lines expressing the antisense transcript fused with the reporter in the genomic context *pasDOG1-LUC::DOG1* (hereafter referred as genAS) reveal a clear increase of the luminescence levels in seedlings treated with CPT after 48h (**Fig. 8**). These results are in agreement with the obtained RT-qPCR results (**Fig. 5**), and support the overexpression of *asDOG1* in response to CPT.

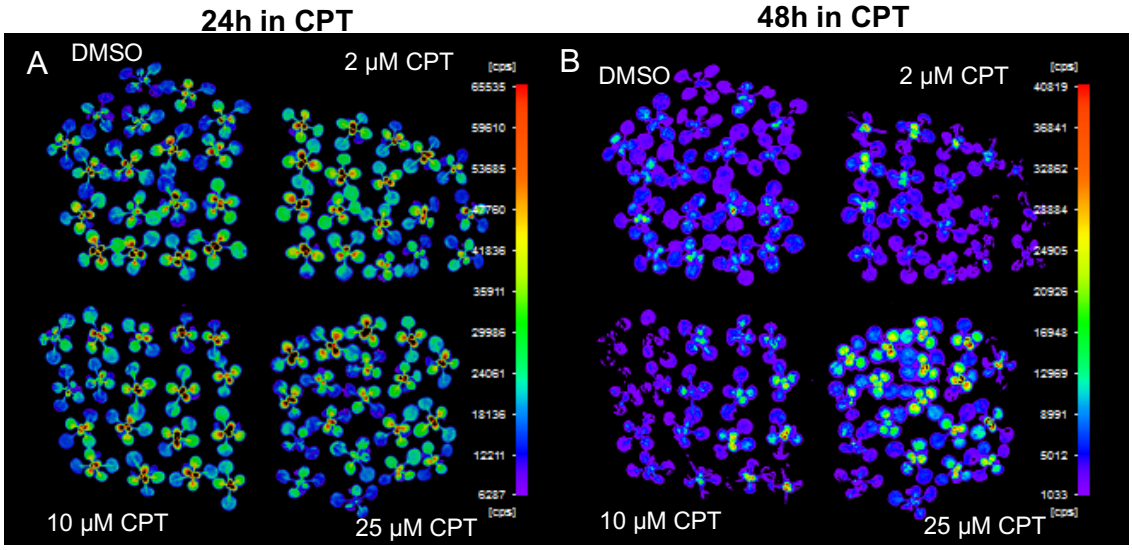


Fig. 8 – ***asDOG1* expression changes in response to CPT in genAS plants**. Representative picture of 12-days old Col-0 seedlings carrying *pasDOG1-LUC::DOG1* (genAS) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 μM (upper right), 10 μM (lower left) and 25 μM of CPT (lower right). *asDOG1* expression is not changed in CPT-treated seedlings after 24h (A) but is strongly increased in seedlings treated with 25 μM of CPT after 48h (B). Heat scale bar represents values of luminescence as counts per second.

TOP1 is responsible for releasing the positive and negative supercoiling during transcription. CPT inhibits TOP1 activity leading to augmented tension ahead of PolII and more relaxed DNA helix behind it. These consequences are predicted to directly impair transcription elongation as reported for human cells (Collins *et al.*, 2001). However, it was also shown that the expression levels are increased upon CPT treatment for several genes (Collins *et al.*, 2001). The authors discussed various explanations for such results, including differences in the distribution and transmission of the tension, and CPT inducing DNA damage indirectly modifying gene expression through the activation of particular signal transduction pathways regulating DNA repair (Collins *et al.*, 2001). Currently, it is known that CPT induces R-loops formation (El Hage *et al.*, 2010; Powell *et al.*, 2013; Groh *et al.*, 2014; Marinello *et al.*, 2016; Shafiq *et al.*, 2017), and such effect can explain an indirect effect of CPT mediated by R-loops that induces gene expression at particular *loci*. Moreover, it is worthy to consider that *DOG1 locus* has two convergent promoters. The movement of two transcriptional machineries toward each other amplifies the positive supercoiling between them (Pannunzio and Lieber, 2016*a*,*b*), making CPT an even bigger threat to gene expression from *DOG1 loci*. Taking this into consideration it was hypothesized that the increased expression levels of *DOG1* and *asDOG1* seen (**Fig. 5 and 8**) may come indirectly from the changes in the DNA topology caused by the CPT treatment rather than a direct consequence of increased DNA supercoiling on *DOG1 loci*, what would be expected to impair transcription and reduce the expression levels.

In summary, the obtained results indicate that the levels of both sense and antisense *DOG1* transcripts are increased after CPT treatment (**Fig. 5 and 8**). However, the luciferase reporter assay for genSense plants (**Fig. 7**) showed contrasting results for *DOG1* expression in response to CPT. Nevertheless, this is suggested to result from the insertion of the luciferase reporter DNA sequence in the construct.

R-loops mediating the CPT effect

TOP1 inhibition by CPT leads to an over-accumulation of negative supercoiling generated during transcription that promotes the formation of R-loops, however, it can have other impacts over transcription independently from R-loops, such as trapping the TOP1 cleavage complexes on the DNA, creating a roadblock for PolII transcription elongation; inducing double-stranded DNA breaks; or simply hindering PolII elongation due to the over-accumulated positive supercoiling (Pommier, 2006; El Hage *et al.*, 2010). Since *pasDOG1::LUC* (hereafter referred as pAS) construct does not include the

genomic region upstream the *DOG1* exon 2, where the two R-loops-forming regions are (*DOG1* promoter and intron 1) according to the ssDRIP-Seq data (**Fig. 3**; Xu et al., 2017), the seedlings carrying this construct were used to evaluate the effect of CPT on *asDOG1* expression in the absence of the putative R-loop-forming regions. In contrast to the effect of this chemical on genAS seedlings, the obtained results showed a clear decrease in the luminescence levels of pAS after 48h on CPT (**Fig. 9**). These results are in agreement with the speculation that the CPT effect on *asDOG1* expression may be mediated through the R-loops formation. As so, it is possible that in pAS seedlings, due to the absence of the R-loops-forming regions, on one hand the CPT no longer is able to induce *asDOG1* expression, and on the other hand the CPT inhibition of TOP1 can negatively influence transcription of the pAS transgene. This would explain the reduced expression observed in pAS seedlings treated with CPT (**Fig. 9**). In fact, the inhibition of TOP1 by CPT in humans revealed that although the CPT effect on gene expression depends on each gene, it generally causes the polymerases to stall during elongation, probably when enough tension is generated ahead of the transcription machineries (Collins *et al.*, 2001), which in the absence of R-loops could result in the repression of *asDOG1* expression as observed in this report (**Fig. 9**).

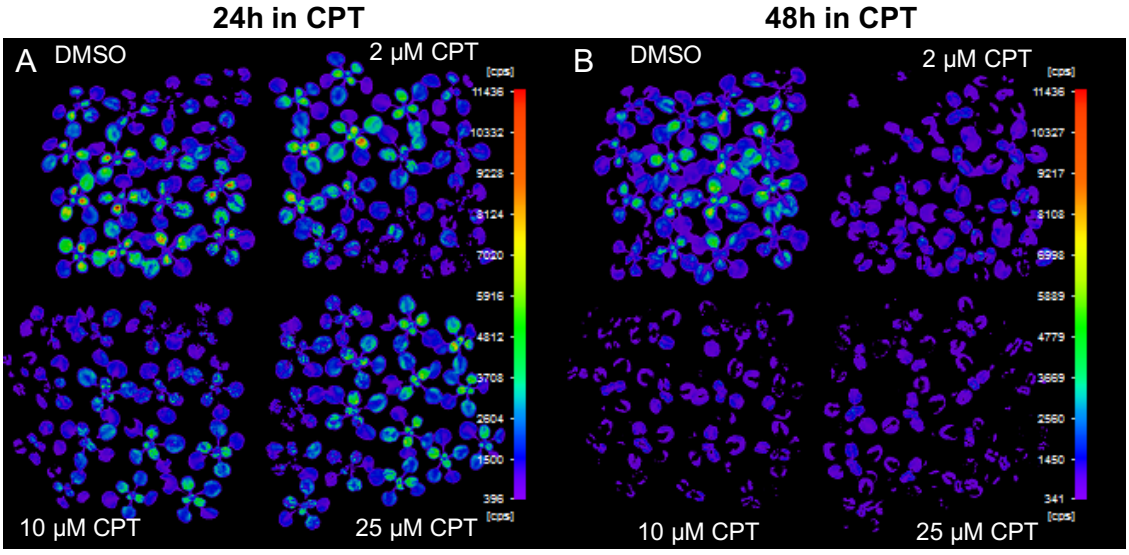


Fig. 9 – ***asDOG1* expression changes in response to CPT in pAS plants**. Representative picture of 12-days old Col-0 seedlings carrying *pasDOG1::LUC* (pAS) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 µM (upper right), 10 µM (lower left) and 25 µM of CPT (lower right). *asDOG1* expression is not changed in CPT-treated seedlings after 24h (A) but is strongly decreased in seedlings treated with 10 and 25 µM of CPT after 48h. Heat scale bar represents values of luminescence as counts per second.

The obtained results showed upregulation of sense and antisense *DOG1* upon CPT treatment. This suggests an indirect effect of CPT treatment rather than the direct changes in DNA supercoiling. Since pAS lacks the genomic region containing the R-loops-forming regions, and based on the fact that *asDOG1* expression is no longer increased in pAS seedlings after CPT (**Fig. 9**) as seen in the genomic context (**Fig. 5 and 8**), it is suggested that R-loops may be the elements mediating the indirect CPT-induction of expression levels at *DOG1 locus*.

Moreover, since *DOG1* represses *asDOG1* and vice versa, an opposite change in expression of sense and antisense *DOG1* was expected as was seen before in different mutants and growth conditions (Fedak *et al.*, 2016; Kowalczyk *et al.*, 2017; Yatusevich *et al.*, 2017). Interestingly, RT-qPCR results showed that both sense and antisense expressions can be induced together by the treatment with CPT. This can be interpreted as if the CPT treatment would influence the ability of the sense and antisense *DOG1* to repress each other, allowing the expression levels of *DOG1* and *asDOG1* to increase simultaneously.

In summary, the results from the reporter assay for genAS and pAS showed that the region upstream the *DOG1* exon 2 is required for the CPT-mediated increase of *asDOG1* expression (**Fig. 8 and 9**).

asDOG1 expression mediating CPT induction of *DOG1*

To assess the role of *asDOG1* in the CPT-mediated increase of *DOG1* expression the luminescence intensity in *psDOG1::LUC* (hereafter referred as pSense) plants after CPT treatment was measured. Strikingly, pSense seedlings growing on CPT for 48h display decreased luminescence levels compared to control (**Fig. 10**). The decrease of luminescence in pSense plants suggests that the antisense transcription at *DOG1 locus* is required for the CPT-mediated increase of sense expression. A similar example in humans was reported for the *VIM locus*, where the transcription of an antisense lncRNA transcript leads to the formation of an R-loop in the promoter region of the *VIM* gene that positively regulates *VIM* sense expression (Boque-Sastre *et al.*, 2015). It is possible that CPT, by inducing changes in the DNA supercoiling, promotes the formation of R-loops within the *DOG1 loci* by the annealing of the *asDOG1* transcripts to the complementary DNA strand, ultimately causing an induction of *DOG1* expression. This *asDOG1*-mediated induction of *DOG1* may not come from a direct positive effect but rather from

the inhibition of the repressive function of *asDOG1*; in other words, the putative R-loop formation by *asDOG1* may block its ability to repress *DOG1*.

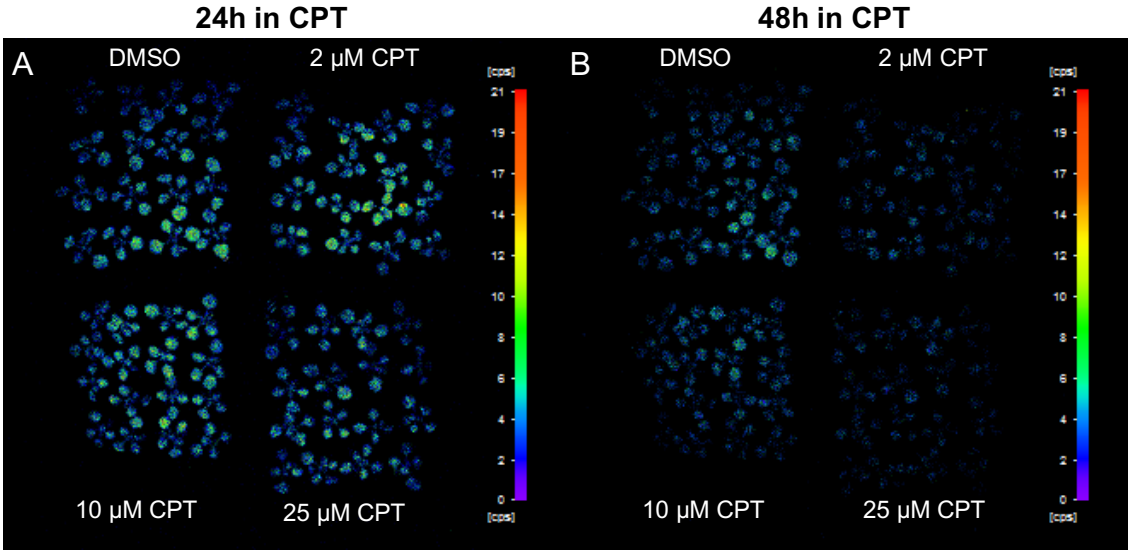


Fig. 10 – ***DOG1* expression changes in response to CPT in pSense plants**. Representative picture of 12-days old Col-0 seedlings carrying *psDOG1::LUC* (pSense) transgene. Seedlings were treated for 24h (A) and 48h (B) with DMSO (mock; upper left), 2 µM (upper right), 10 µM (lower left) and 25 µM of CPT (lower right). *DOG1* expression is not changed in CPT-treated seedlings after 24h (A) but is slightly decreased in seedlings treated with 25 µM of CPT after 48h. Heat scale bar represents values of luminescence as counts per second.

In summary, the reporter assay performed with pSense seedlings revealed a downregulation of *DOG1* after CPT treatment (**Fig. 10**). Based on the previous results, it is suggested that *asDOG1* expression mediates the CPT increase of *DOG1* expression.

Detection of R-loops formation on *DOG1 loci*

To further clarify the role of R-loops in *DOG1* regulation, a technique called DNA-RNA Immunoprecipitation (DRIP) was established in the laboratory as a consequence of this study. DRIP is a variant of ChIP (Chromatin Immunoprecipitation) that applies the monoclonal antibody S9.6 which recognizes DNA-RNA hybrids (Hu *et al.*, 2006). Briefly, nuclei of plant material were isolated and used to extract genomic DNA that was then treated with RNase A and Proteinase K to degrade ssRNA and proteins respectively, and sonicated prior immunoprecipitation (IP). After IP with the S9.6 antibody, the DNA-RNA hybrids are captured with magnetic beads and washed extensively. The eluted hybrids are then used for qPCR (DRIP-qPCR; **Fig. 11**). DNA from each biological replicate used for IP was in parallel digested with RNase H as a negative control, since

RNase H degrades RNA hybridized to DNA. “No antibody” was used as a control for the IP itself, in which water was added to the DNA instead of the S9.6 antibody. From all samples 10% of the volume was taken just before IP, diluted in the same final volume as the eluted samples after IP and used for qPCR as input allowing the calculation of the percentage of DNA used for IP that was eluted (named “percent of input”).

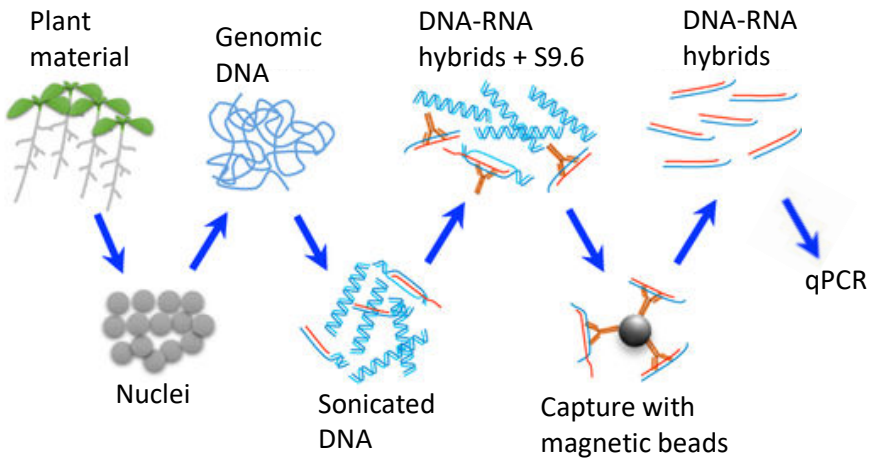


Fig. 11 – **Framework of the DRIP-qPCR procedure optimised in this work** (adapted from Xu et al., 2017). Plant material was collected and used for nuclei isolation without any crosslinking step (opposite to standard ChIP), then genomic DNA was extracted and sonicated, and used for IP with the S9.6 monoclonal antibody. Next, hybrids were isolated using magnetic beads, and eluted for further qPCR analysis.

To select reliable and efficient primer pairs a qPCR was performed on 10% input with several primer pairs available in the laboratory and newly designed primers; the primer pairs selected had low Ct values between 21 and 22°C (**Table S2**). Melting curves on qPCR confirmed the amplification of one single product that was ran on an agarose gel to confirm its size.

DRIP-qPCR was first performed on adult Col-0 plant leaves. DRIP-qPCR results were plotted as percent of input (**Fig. 12**). The obtained results revealed the enrichment of immunoprecipitated DNA-RNA hybrids over the *DOG1* promoter and exon2-intron2 junction comparing to low enriched regions at the end of the *DOG1* gene. These results are in agreement with the formation of an R-loop on the promoter region of *DOG1* as shown by the ssDRIP-Seq data (**Fig. 3**; Xu et al., 2017), however, no strong enrichment of R-loops was detected within the *DOG1* intron 1. Co-transcriptionally formed R-loops were shown to be strongly influenced by the expression levels (Ginno *et al.*, 2012; Chen *et al.*, 2015; Sanz *et al.*, 2016; Wahba *et al.*, 2016). Since *DOG1* and *asDOG1* expression levels strongly vary during development, and in response to different growth conditions

such as water, light, temperature and sugars concentration on the media, the differences in culture conditions between independent studies may justify the differences in the enrichment of the observed R-loops within *DOG1* intron 1. It is also possible that the differences in the R-loops formation pattern detected through DRIP-qPCR in this study may be due to the use of a different approach. These are all plausible explanations for the different enrichment of R-loops over the intron 1 between the DRIP-qPCR results and the ssDRIP-Seq data (**Fig. 3**; Xu et al., 2017).

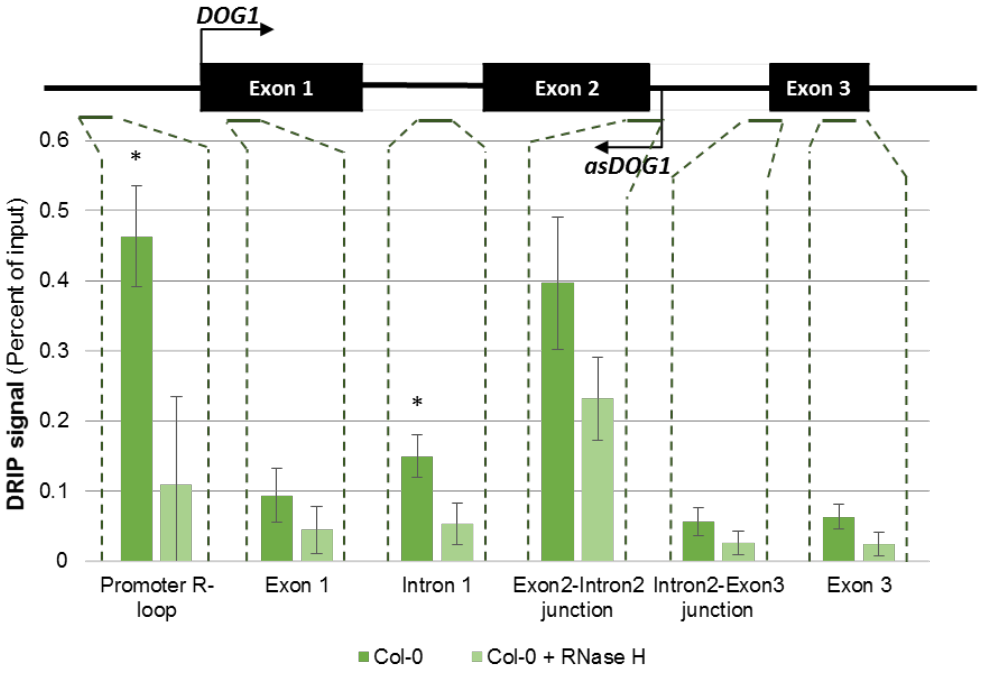


Fig. 12 – **R-loops detection within *DOG1 loci* in adult plants**. DRIP-qPCR on leaves of adult Col-0 plants with the selected primers for *DOG1* (*At5G45830.1*). Results shown as percent of input for samples not treated (Col-0) and treated with 7.5 U of recombinant *E. coli* RNase H (NEB, M0297S) overnight at 37°C (Col-0 + RNase H) as negative control. Strong signal is detected over the *DOG1* promoter region and exon2-intron2 junction. RNase H treatment prior IP decreased the signal. Bars show the average for three biological replicates, and error bars show the standard deviation. * show significant differences between the treated and not treated samples for each region of *DOG1*, with *t*-test for $p < 0.05$. On top is the schematic representation of DOG1 locus with the amplified region marked with green lines. Dashed lines match the amplified regions with the corresponding bars in the plot.

Surprisingly, the obtained DRIP-qPCR results also shown a strong signal over the exon2-intron2 junction (**Fig. 12**). Yet, the signal is not strongly decreased after RNase H treatment, suggesting that the signal detected from qPCR either comes from an RNase H-resistant DNA-RNA hybrid or does not come from a DNA-RNA hybrid. A recent study revealed that the S9.6 antibodies used in this work for DRIP also have affinity for dsRNAs interfering with the R-loops detection (Hartono *et al.*, 2018). Since the antibody

recognizes dsRNAs, and the exon2-intron2 junction is close to *asDOG1* TSS, the signal detected by qPCR may come from a RNA duplex formed by the annealing of sense and antisense transcripts or from a RNA secondary structure of one of the two transcripts. Nevertheless, since there was no reverse transcription step after IP, the amplification by qPCR should not work if the template was RNA, unless the dsRNA was immunoprecipitated together with DNA. One way to address this question would be including an RNase III treatment prior IP to degrade the dsRNAs. Incorporating this step on the experimental procedure used in this study should abolish the signal detected by qPCR for the samples not treated with RNase H if the signal detected (**Fig. 12**) comes from a dsRNA-DNA structure, but should remain unchanged if the signal comes from an RNase H-resistant DNA-RNA hybrid. Additionally, to test if the signal comes from a dsRNA trapped in the chromatin as hypothesized, it would also be relevant to use an antibody that specifically recognizes dsRNAs (Schönborn *et al.*, 1991) instead of the S9.6 antibody. Since *asDOG1* was shown to work in *cis* but not in *trans* (Fedak et al., 2016), the previous observations led us to speculate that the formation of a sense-antisense dsRNA would possibly be involved in the currently puzzling molecular mechanism by which *asDOG1* represses *DOG1* in *cis*.

Despite the strong decrease of immunoprecipitated R-loops in samples treated with RNase H (**Fig. 12**), based on preliminary observations during the initial steps of optimization of the protocol, it is possible that RNase H treatment may have not been sufficient to degrade all DNA-RNA hybrids. For the previous experiment (**Fig. 12**), 20 µg of DNA were used for IP and 7.5 U of RNase H were used for the treatment. Next, DRIP-qPCR was performed using material from 10-days old Col-0 seedlings grown on plates. For IP 15 µg of DNA were used, and for the RNase H treatment 15 U of enzyme were used. According to the previous results, a strong enrichment of hybrids over the *DOG1* promoter region comparing with the rest of *DOG1 locus* (**Fig. 13**) was detected. The increase in the amount of enzyme used for the RNase H treatment resulted in a stronger decrease of signal from the treated samples, subsequently revealing significant differences between the treated and not treated samples for all regions of *DOG1 locus* except for exon2-intron2 junction. This suggests that RNase H concentration was no longer limiting the reaction. Thus, since the hybrids enrichment over exon2-intron2 junction was not significantly decreased after the RNase H treatment, similarly to the previous experiment, the idea that the signal detected by qPCR does not come from a conventional R-loop but may come from a more complex structure, possibly an RNA duplex stacked in the chromatin is further reinforced.

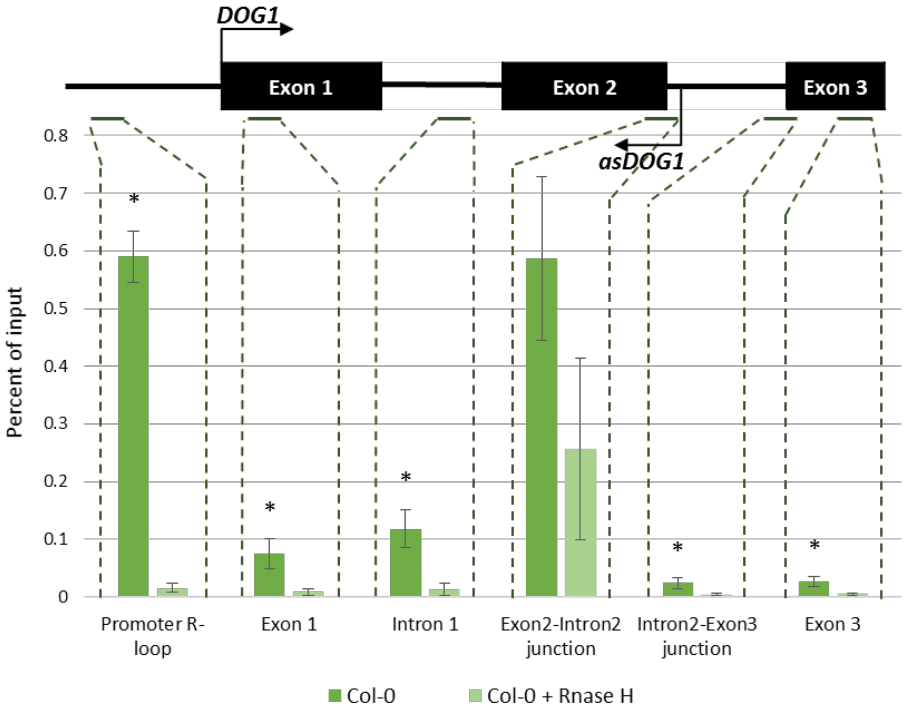


Fig. 13 – **R-loops detection within *DOG1 loci* in young seedlings**. DRIP-qPCR on 10-days old Col-0 seedlings with the selected primers for *DOG1* (*At5G45830.1*). Results show the percent of input for samples not treated (Col-0) and treated with 15 U of recombinant *E. coli* RNase H (NEB, M0297S) overnight at 37°C (Col-0 + RNase H) as negative control. RNase H treatment prior IP decreased the signal as expected. Strong signal is detected over the *DOG1* promoter region and exon2-intron2 junction. Bars show the average for three biological replicates, and error bars show the standard deviation. * show significant differences between the treated and not treated samples for each region of *DOG1*, with *t*-test for $p < 0.05$. On top is the schematic representation of DOG1 locus with the amplified region marked with green lines. Dashed lines match the amplified regions with the corresponding bars in the plot.

Overall, DRIP-qPCR results remarkably showed a strong enrichment of R-loops formed over the *DOG1* promoter region. This was previously identified by ssDRIP-Seq on a genome-wide study (**Fig. 3**; Xu et al., 2017) thus supporting the presented results. The identification of a R-loop formed on the promoter of *DOG1* by DRIP-qPCR together with the results from RT-qPCR and the reporter assay support the hypothesis that the CPT effect over *DOG1* and *asDOG1* expression may indeed be mediated by changes in the R-loops formation.

Interestingly, the obtained results revealed that the expression of both *DOG1* sense and antisense is increased upon CPT treatment. This can be interpreted as if the mutual exclusive pattern of the sense and antisense pair seen before (Fedak *et al.*, 2016) is disrupted after CPT treatment. It was hypothesized that the increased R-loops formation in response to CPT blocked the ability of the sense and antisense pair to repress each other. Previously published results showed that the transcription

termination site (TTS) selection of the sense transcript and deposition of H2Bub are possibly involved in the mechanism of *DOG1*-mediated repression of *asDOG1* (Kowalczyk *et al.*, 2017), although the molecular mechanism by which *asDOG1* represses its sense counterpart remains unknown.

A plausible model explaining the obtained results is that the *asDOG1* transcript forms an R-loop at the *DOG1* sense promoter that acts as a traffic light for the transcriptional initiation, controlling transcription bursts from the sense promoter, allowing a more coordinated transcriptional process (**Fig. 14A and B**). In other words, DOG1 sense transcription blocks asDOG1 transcription initiation events through PolII readthrough over its promoter (**Fig. 14A**). After a *DOG1* burst, *asDOG1* transcription starts and it leads to an R-loop formation at the *DOG1* promoter that can affect the binding of transcription factors, or chromatin conformation over that region, and transiently switch off transcription of *DOG1* sense, thus avoiding conflicts such as PolII collisions, RNA duplexes formation between the sense and antisense transcripts, competition for RNA binding proteins and other processing factors, etc. (**Fig. 14B**). Afterwards, once the R-loop is resolved by a specific cellular machinery, *DOG1* sense transcription is triggered and its readthrough over the *asDOG1* promoter stops the antisense transcription initiation again (**Fig. 14A**). Overall, this orchestrated transcription mediated by R-loops would inhibit the sense-antisense mutual repression resulting in the optimal transcriptional levels from sense and antisense at the same *locus*. It is hypothesized that CPT induces the R-loops formation at *DOG1* promoter more frequently, thus keeping the synchronized transcription and enhancing sense and antisense expression levels (**Fig. 14B**). This model is consistent with the recent findings that transcriptional bursting is a general property of transcription that influences gene expression in various organisms (Nicolas *et al.*, 2017). Additionally, this mechanism would allow to keep *asDOG1* expression levels high enough to mediate *DOG1* repression rapidly after the right signaling stimuli, which is an expensive strategy, although often used by plants. This is in agreement with the observations that *asDOG1* expression increases during seed maturation along with the increase of sense expression without any apparent mutual repression. Only at a particular stage of development *asDOG1* expression keeps high and *DOG1* is dramatically downregulated (Fedak *et al.*, 2016). According to the proposed model it would be expected that at that stage a trigger would activate or direct the degradation of the R-loops formed at *DOG1* promoter, and the high transcription no longer coordinated would generate conflicts that underlay/activate the mutual repression.

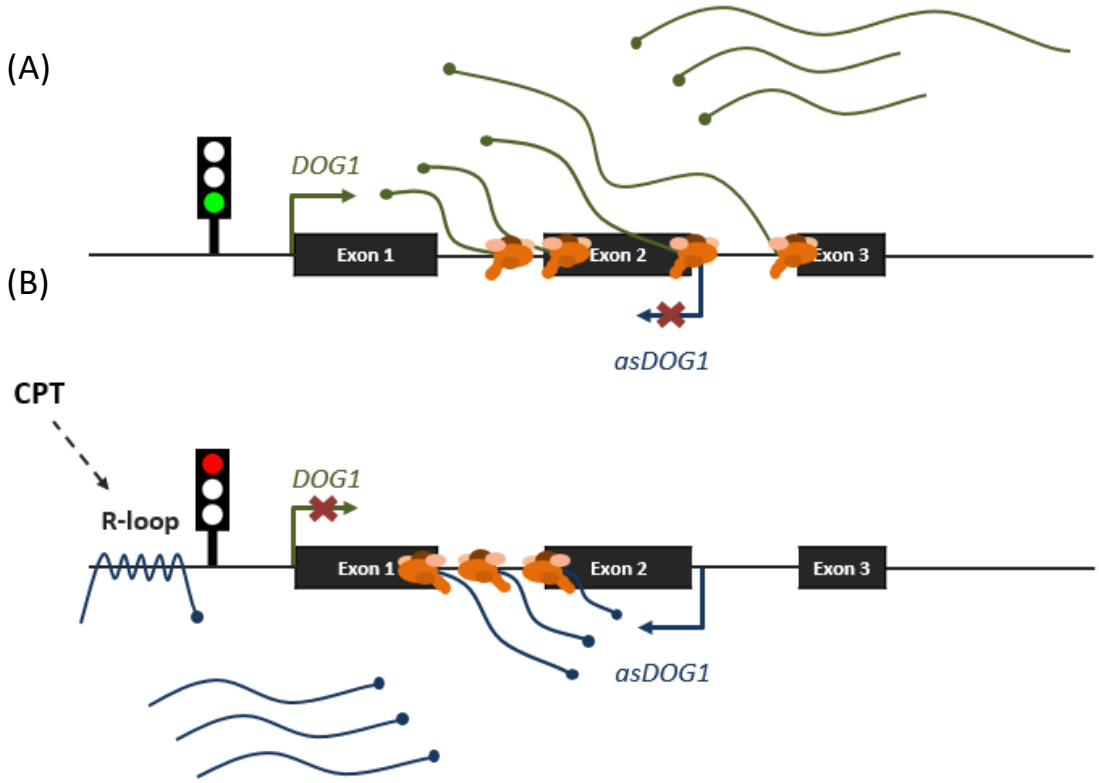


Fig. 14 – Model of R-loops assisting in the coordination of transcription bursts from sense and antisense *DOG1* promoters. (A) Transcription bursts from *DOG1* sense promoter lead to the transcription of *shDOG1* and *lgDOG1* transcript isoforms. Transcription of *lgDOG1* results in the readthrough of *asDOG1* promoter and is thought to mediate the repression of *asDOG1* transcription initiation. (B) After the complete round of sense transcription, asDOG1 promoter is susceptible to be activated. Antisense transcripts possibly derived from *asDOG1* transcription form an R-loop over the sense promoter region which shut down *DOG1* transcription initiation and allows *DOG1* to bypass downstream *asDOG1*-mediated repression events. Once the R-loop is resolved by a specialized cellular machinery, sense transcription can be resumed. In the presence of CPT, R-loops formation is thought to increase, leading to a more frequent orchestration of sense and antisense transcription bursts. In this condition, the optimal sense and antisense transcription is achieved. Transcription from the antisense promoter during sense transcription events would lead to conflicts such as PolII collisions, dsRNAs formation between the transcripts, competition for RNA binding proteins and other processing factors, etc.

At the *FLC locus* in Arabidopsis, a recent study addressed the mutual repression between *FLC* and its antisense transcript *COOLAIR* using single-molecule RNA FISH. This technique allowed the authors to observe single RNA molecules at the single cell level. Remarkably, they found that although the sense and antisense transcripts can co-occur in the same cell, they are mutually exclusive at individual *loci* (Rosa *et al.*, 2016). For *DOG1* there is no information so far regarding the transcription from individual *loci*. Since *asDOG1* was seen to act in *cis* but not in *trans* (Fedak *et al.*, 2016), transcription of both sense and antisense *DOG1* from the same *loci* is expected. However, in the case

of failure of capturing both transcripts at the same *locus* that would not necessarily mean that their transcription cannot co-occur in single *loci*. Coordinated transcriptional bursts from sense and antisense promoters, particularly in such short *locus* (around 2 kb) such as *DOG1*, would probably not allow to clearly visualize both sense and antisense transcriptional events at the same *locus* at the same timepoint by smFISH.

Concluding Remarks

Together, the RT-qPCR results and the luciferase reporter assay for genAS showed that upon CPT treatment both sense and antisense *DOG1* expression was increased (**Fig. 5 and 8**). Moreover, it was demonstrated that the CPT-mediated increase of *asDOG1* expression requires the region upstream *DOG1* exon 2 (**Fig. 8 and 9**), where DRIP-qPCR revealed the formation of an R-loop according with the ssDRIP-Seq data (**Fig. 3**; Xu et al., 2017). This suggests that the CPT effect on *asDOG1* expression may work through the R-loops formation. Finally, the results from the reporter assay for genSense and pSense showed that the CPT-mediated increase of *DOG1* sense expression requires the genomic region downstream exon 2 (**Fig. 7 and 10**). This observation suggests that in the absence of *asDOG1* expression, CPT no longer affects *DOG1* expression. Altogether, and since the promoter R-loop was shown to be made by an antisense transcript (Xu et al., 2017), it is speculated that an *asDOG1* long transcript isoform or a readthrough after 3' end cleavage of *asDOG1* transcript forms the R-loop at *DOG1* promoter that, in turn, mediates overexpression of both sense and antisense *DOG1* when its formation is promoted by CPT. Due to the mutual repressive behaviour of the sense and antisense pair at *DOG1*, this work strongly allows to propose a model in which R-loops formation at *DOG1* promoter make the *DOG1* sense promoter immunue to *asDOG1*-mediated repression. This model assumes synchronized sense and antisense transcription. Although this hypothesis lacks further experimental validation, the obtained results provide interesting clues on a novel mechanism of regulation of *DOG1* gene expression, and raised important questions that need to be addressed in the future. Does CPT induce increased R-loops formation at *DOG1* promoter? Is the R-loop formed as a result of transcription of the previously found antisense transcript (*asDOG1*; Fedak *et al.*, 2016)? Is the R-loops formation *per se* causing the changes in sense and antisense expression profiles? Is this regulatory process playing a relevant role in plants' physiology?

Future Perspectives

In this study, R-loops at *DOG1* promoter were detected, and changes in sense and antisense *DOG1* expression were recorded in response to a chemical that was shown before to induce R-loops formation. However, CPT inhibition of TOP1 affects the transcription of virtually all the nascent transcripts during elongation and can influence gene expression independently from R-loops formation. The validation of the ssDRIP-Seq data (**Fig. 3**; Xu et al., 2017) confirming the presence of R-loops at the *DOG1* promoter (**Fig. 12 and 13**) does not signify that the CPT-induced changes on *DOG1* and *asDOG1* expression levels are mediated by the R-loops formation. Thus, the next logical step would be to directly test the R-loops formation pattern upon CPT treatment by DRIP-qPCR. This would allow to conceptually link the CPT effect on *DOG1* and *asDOG1* expression to changes in R-loops formation.

DRIP-qPCR was successfully established in the laboratory as a consequence of this study. Although it allowed the detection of R-loops enrichment over the *DOG1* genomic region, this technique has its limitations such as not allowing a strand-specific detection. In that sense, a diversity of DRIP variants was recently developed mostly relying on sequencing (Vanoosthuyse, 2018). This limitation made it impossible in this work to determine if the transcript forming the R-loop was a sense or an antisense transcript. The CPT-mediated effect on *DOG1* sense expression was shown to be dependent on *asDOG1* expression by the luciferase reporter assays (**Fig. 7 and 10**) making *asDOG1* transcript the first suspect. This is in agreement with the ssDRIP-Seq data, showing that the R-loop on the promoter of *DOG1* is formed by an antisense transcript (**Fig. 3**; Xu et al., 2017). Yet, is intriguing whether the RNA forming the R-loop at the *DOG1* promoter is the *asDOG1* described by Fedak *et al.* (2016) or a different antisense ncRNA. Results from the reanalysis of strand-specific direct RNA sequencing (DRS)-based mapping of polyadenylation sites in the Arabidopsis genome (Sherstnev *et al.*, 2012) showed two prominent TTSs for *asDOG1*: one within the intron 1 and the other near *DOG1* TSS **(Fig. S3**; Fedak et al., 2016). These observations suggest that the transcription of *asDOG1* transcripts should mainly terminate before the R-loop-forming region on *DOG1* promoter. It is still possible that a fraction of *asDOG1* transcripts fail to terminate transcription on the mentioned TTSs forming the R-loop on *DOG1* promoter, or the R-loop is formed by the PolII readthrough transcript after 3' end cleavage. However, an interesting possibility is that the R-loop is formed by a different unannotated ncRNA. One way to identify the orientation of the RNA responsible for forming the R-

loop on *DOG1* promoter could be using strand-specific reverse transcription-based methods to map the putative ncRNA transcript in the R-loop (**Fig. S4**). Reverse transcription using the forward primer would lead to the synthesis of cDNA from an antisense transcript while reverse transcription using the reverse primer would lead to the synthesis of a cDNA from a sense transcript. Following reverse transcription, the samples could be used for PCR with both primers allowing the identification of the orientation (sense or antisense) of the RNA forming the R-loop (**Fig. S4**). An additional experiment that could be done to understand if the RNA forming the R-loop is phisicaly connected to *asDOG1* (if it is an *asDOG1* transcript or a readthrough transcript derived from *asDOG1*) would be to perform DRIP-qPCR on *dog1-5* (SALK_022748) mutant that carries a T-DNA insertion within *DOG1* exon 3. This T-DNA insertion is thought to disrupt an important *cis*-regulatory element controlling *asDOG1* expression, since it was shown that in seeds of this mutants *asDOG1* is strongly downregulated (Fedak *et al.*, 2016). However, in seedlings or older plants *asDOG1* is not downregulated in *dog1-5* (unpublished data). Therfore the use of *dog1-5* would require optimization of the herein described DRIP protocol (mostly on the nuclei isolation step) to perform this experiment on seeds. If the R-loop is formed by a transcript resulting from *asDOG1* transcription, its formation should be abolished on *dog1-5* mutants. Otherwise this would support the existence of a novel ncRNA at *DOG1 locus* or the formation of the R-loop by a trans-acting RNA.

Additionally, it would also be interesting to assess whether the CPT effect over sense and antisense expression is achieved through the blocking of the mutual exclusive repression function on *DOG1 locus*. Our laboratory recently found that the *hub1-5* mutant (lacking the H2Bubq-depositing enzyme HUB1) displays increased *asDOG1* and decreased *DOG1* expression comparing to the wild type (Kowalczyk *et al.*, 2017). Thus, it is speculated that if CPT blocks the mutual repression on *DOG1 locus*, then after the CPT treatment of *hub1-5*, the sense *DOG1* expression should no longer be downregulated comparing to the wild type. That would support the idea that at least *asDOG1* loses the ability to repress *DOG1* after CPT treatment. On the other way around, using the same approach, CPT could be used to block the *DOG1*-repression of *asDOG1*. For this, it could eventually be possible to use the *fy-2* mutant characterized in this laboratory to display upregulation of the *lgDOG1* isoform, suggested to be involved in the *asDOG1* repression through readthrough over *asDOG1* promoter (Kowalczyk *et al.*, 2017).

Another future goal would be to address the biological significance of the R-loops formation on *DOG1*. *DOG1* is the key *locus* controlling the release of seed dormancy in *A. thaliana* (Alonso-Blanco *et al.*, 2003; Bentsink *et al.*, 2006; Fedak *et al.*, 2016). Additionally, this lab previously showed that *asDOG1* plays an important role in the perception of environmental cues in plants subjected to drought (Yatusevich *et al.*, 2017). To address the biological relevance of the R-loops formation within the *DOG1 loci*, DRIP-qPCR could be used to determine if the R-loops formation is changed in seeds during the maturation stages when the dynamics between *DOG1* and *asDOG1* were seen to play the main role controlling seed dormancy (Fedak *et al.*, 2016), and in leaves of adult plants subjected to water deprivation when the changes on *DOG1* and *asDOG1* expression levels were shown to affect plants ability to recover from drought stress (Yatusevich *et al.*, 2017).

Overall the present study shed light on the function of R-loops in the regulation of gene expression in plants, and generated remarkable insights over the currently unknown molecular mechanism by which *asDOG1* regulates *DOG1* expression. In addition to the known regulatory pathways that act to fine tune *DOG1* expression, this study provided evidences that this gene may also be regulated by R-loops formation, potentially contributing to the intricate regulation of seed dormancy/germination and the response to drought stress in adult plants. So far, the only example of a functional R-loop in *A. thaliana* nuclear genome was found to integrate the sense and antisense pair regulation at *FLC locus* (Sun et al., 2013). The recent mapping of R-loops in the Arabidopsis genome found a remarkable prevalence of antisense R-loops being predominantly formed over the TSS of the sense gene (Xu et al., 2017). Is the antisense R-loops formation part of a more general mechanism of regulation of transcription initiation in Arabidopsis?

References

Aguilera A, García-Muse T. 2012. R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular Cell **46**, 115–124.

Alonso-Blanco C, Bentsink L, Hanhart CJ, Vries HB, Koornneef M. 2003. Analysis of Natural Allelic Variation at Seed Dormancy *Loci* of *Arabidopsis thaliana*. Genetics **164**, 711–729.

Bentsink L, Jowett J, Hanhart CJ, Koornneef M. 2006. Cloning of *DOG1*, a quantitative trait *locus* controlling seed dormancy in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **103**, 17042–7.

Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature **511**, 362–365.

Bonnet A, Grosso AR, Elkaoutari A, Coleno E, Presle A, Sridhara SC, Janbon G, Géli V, de Almeida SF, Palancade B. 2017. Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Molecular Cell **67**, 608–621.e6.

Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S. 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America **112**, 5785–90.

Boule J-B, Zakian VA. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic Acids Research **35**, 5809–5818.

Chen PB, Chen H V, Acharya D, Rando OJ, Fazzio TG. 2015. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nature Structural & Molecular Biology **22**, 999–1007.

Collins I, Weber A, Levens D. 2001. Transcriptional consequences of topoisomerase inhibition. Molecular and Cellular Biology **21**, 8437–51.

Conn VM, Hugouvieux V, Nayak A, *et al.* 2017. A circRNA from *SEPALLATA3* regulates splicing of its cognate mRNA through R-loop formation. Nature Plants **3**, 17053.

Cristini A, Groh M, Kristiansen MS, Gromak N. 2018. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Reports **23**, 1891–1905.

Cyrek M, Fedak H, Ciesielski A, *et al.* 2016. Seed Dormancy in *Arabidopsis* Is Controlled by Alternative Polyadenylation of *DOG1*. Plant Physiology **170**, 947–55.

Dinh TT, Gao L, Liu X, *et al.* 2014. DNA Topoisomerase 1α Promotes Transcriptional Silencing of Transposable Elements through DNA Methylation and Histone Lysine 9 Dimethylation in Arabidopsis (O Voinnet, Ed.). PLOS Genetics **10**, e1004446.

Dolata J, Guo Y, Kołowerzo A, Smoliński D, Brzyżek G, Jarmołowski A, Świeżewski S. 2015. NTR1 is required for transcription elongation checkpoints at alternative exons in *Arabidopsis*. The EMBO journal **34**, 544–58.

Drolet M, Phoenix P, Menzel R, Massé E, Liu LF, Crouch RJ. 1995. Overexpression of RNase H partially complements the growth defect of an *Escherichia coli delta topA* mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proceedings of the National Academy of Sciences of the United States of America **92**, 3526–30.

Fedak H, Palusinska M, Krzyczmonik K, Brzezniak L, Yatusevich R, Pietras Z, Kaczanowski S, Swiezewski S. 2016. Control of seed dormancy in *Arabidopsis* by a *cis*-acting noncoding antisense transcript. Proceedings of the National Academy of Sciences of the United States of America **113**, E7846–E7855.

Finch-Savage WE, Leubner-Metzger G. 2006. Seed dormancy and the control of germination. New Phytologist **171**, 501–523.

Footitt S, Müller K, Kermode AR, Finch-Savage WE. 2015. Seed dormancy cycling in *Arabidopsis*: chromatin remodelling and regulation of *DOG1* in response to seasonal environmental signals. The Plant Journal **81**, 413–425.

Frey A, Audran C, Marin E, Sotta B, Marion-Poll A. 1999. Engineering seed dormancy by the modification of zeaxanthin epoxidase gene expression. Plant Molecular Biology **39**, 1267–1274.

Gao M, Wei W, Li M-M, *et al.* 2014. Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. Cell Research **24**, 532–541.

García-Rubio ML, Pérez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado I V., Aguilera A. 2015. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops (J Sekelsky, Ed.). PLOS Genetics **11**, e1005674.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Molecular Cell **45**, 814–825.

Groh M, Lufino MMP, Wade-Martins R, Gromak N. 2014. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome (A Aguilera, Ed.). PLOS Genetics **10**, e1004318.

El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes & Development **24**, 1546–58.

Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, Vanoosthuyse V. 2018. The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. Journal of Molecular Biology **430**, 272–284.

Hatchi E, Skourti-Stathaki K, Ventz S, *et al.* 2015. BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Molecular Cell **57**, 636–647.

Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH. 2006. An antibody-based microarray assay for small RNA detection. Nucleic Acids Research **34**, e52.

Huertas P, Aguilera A. 2003. Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. Molecular Cell **12**, 711–721.

Huo H, Wei S, Bradford KJ. 2016. *DELAY OF GERMINATION1* (*DOG1*) regulates both seed dormancy and flowering time through microRNA pathways. Proceedings of the National Academy of Sciences of the United States of America **113**, E2199-206.

Keskin H, Shen Y, Huang F, Patel M, Yang T, Ashley K, Mazin A V., Storici F. 2014. Transcript-RNA-templated DNA recombination and repair. Nature **515**, 436–439.

Kowalczyk J, Palusinska M, Wroblewska-Swiniarska A, Pietras Z, Szewc L, Dolata J, Jarmolowski A, Swiezewski S. 2017. Alternative Polyadenylation of the Sense

Transcript Controls Antisense Transcription of *DELAY OF GERMINATION 1* in *Arabidopsis*. Molecular Plant **10**, 1349–1352.

Leon-Kloosterziel KM, Gil MA, Ruijs GJ, Jacobsen SE, Olszewski NE, Schwartz SH, Zeevaart JAD, Koornneef M. 1996. Isolation and characterization of abscisic acid-deficient *Arabidopsis* mutants at two new *loci*. The Plant Journal **10**, 655–661.

Li X, Manley JL. 2005. Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. Cell **122**, 365–378.

Liu Y, Koornneef M, Soppe WJJ. 2007. The Absence of Histone H2B Monoubiquitination in the *Arabidopsis hub1* (*rdo4*) Mutant Reveals a Role for Chromatin Remodeling in Seed Dormancy. The Plant Cell Online **19**, 433–444.

Marinello J, Bertoncini S, Aloisi I, Cristini A, Malagoli Tagliazucchi G, Forcato M, Sordet O, Capranico G. 2016. Dynamic Effects of Topoisomerase I Inhibition on R-Loops and Short Transcripts at Active Promoters (F Leng, Ed.). PLOS ONE **11**, e0147053.

Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular Cell **41**, 21–32.

Nakabayashi K, Bartsch M, Xiang Y, Miatton E, Pellengahr S, Yano R, Seo M, Soppe WJJ. 2012. The time required for dormancy release in *Arabidopsis* is determined by *DELAY OF GERMINATION1* protein levels in freshly harvested seeds. The Plant Cell **24**, 2826–38.

Nicolas D, Phillips NE, Naef F. 2017. What shapes eukaryotic transcriptional bursting? Molecular BioSystems **13**, 1280–1290.

Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T. 2016. Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. Cell **167**, 1001–1013.e7.

Pannunzio NR, Lieber MR. 2016*a*. RNA Polymerase Collision versus DNA Structural Distortion: Twists and Turns Can Cause Break Failure. Molecular Cell **62**, 327–334.

Pannunzio NR, Lieber MR. 2016*b*. Dissecting the Roles of Divergent and Convergent Transcription in Chromosome Instability. Cell Reports **14**, 1025–1031.

Pommier Y. 2006. Topoisomerase I inhibitors: Camptothecins and beyond. Nature Reviews Cancer **6**, 789–802.

Powell WT, Coulson RL, Gonzales ML, *et al.* 2013. R-loop formation at *Snord116* mediates topotecan inhibition of *Ube3a*-antisense and allele-specific chromatin decondensation. Proceedings of the National Academy of Sciences **110**, 13938–43.

Roberts RW, Crothers DM. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science **258**, 1463–6.

Rosa S, Duncan S, Dean C. 2016. Mutually exclusive sense–antisense transcription at *FLC* facilitates environmentally induced gene repression. Nature Communications **7**, 13031.

Roy D, Lieber MR. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. Molecular and Cellular Biology **29**, 3124–33.

Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. 2010. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Molecular and Cellular Biology **30**, 146–59.

Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chédin F. 2016. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Molecular Cell **63**, 167–178.

Schönborn J, Oberstrass J, Breyel E, Tittgen J, Schumacher J, Lukacs N. 1991. Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. Nucleic Acids Research **19**, 2993–3000.

Shafiq S, Chen C, Yang J, *et al.* 2017. DNA Topoisomerase 1 Prevents R-loop Accumulation to Modulate Auxin-Regulated Root Development in Rice. Molecular Plant **10**, 821–833.

Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton GJ, Simpson GG. 2012. Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. Nature Structural & Molecular Biology **19**, 845–852.

Shirzadegan M, Christie P, Seemann JR. 1991. An efficient method for isolation of RNA from tissue cultured plant cells. Nucleic Acids Research **19**, 6055.

Shu H, Wildhaber T, Siretskiy A, Gruissem W, Hennig L. 2012. Distinct modes of DNA accessibility in plant chromatin. Nature Communications **3**, 1281.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature **516**, 436–9.

Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & Development **28**, 1384–96.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular Cell **42**, 794–805.

Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. Molecular Cell **56**, 777–785.

Song C, Hotz-Wagenblatt A, Voit R, Grummt I. 2017. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes and Development **31**, 1370–1381.

Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. 2013. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC locus*. Science **340**, 619–621.

Tanikawa M, Sanjiv K, Helleday T, Herr P, Mortusewicz O. 2016. The spliceosome U2 snRNP factors promote genome stability through distinct mechanisms; transcription of repair factors and R-loop processing. Oncogenesis **5**, e280.

Tous C, Aguilera A. 2007. Impairment of transcription elongation by R-loops in vitro. Biochemical and Biophysical Research Communications **360**, 428–432.

Vanoosthuyse V. 2018. Strengths and Weaknesses of the Current Strategies to Map and Characterize R-Loops. Non-Coding RNA **4**, 9.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes & Development **30**, 1327–38.

Wei W, Ba Z, Gao M, Wu Y, Ma Y, Amiard S, White CI, Rendtlew Danielsen JM, Yang Y-G, Qi Y. 2012. A Role for Small RNAs in DNA Double-Strand Break Repair. Cell **149**, 101–112.

Xiong L, Zhu J-K, Endo A, Okamoto M, Koshiba T, Cheng W-H. 2003. Regulation of Abscisic Acid Biosynthesis. Plant Physiology **133**, 29–36.

Xu W, Xu H, Li K, Fan Y, Liu Y, Yang X, Sun Q. 2017. The R-loop is a common chromatin feature of the Arabidopsis genome. Nature Plants **3**, 704–714.

Yatusevich R, Fedak H, Ciesielski A, Krzyczmonik K, Kulik A, Dobrowolska G, Swiezewski S. 2017. Antisense transcription represses *Arabidopsis* seed dormancy QTL *DOG1* to regulate drought tolerance. EMBO Reports **18**, 2186–2196.

Supplemental Information

Table S1 – Table of primers used in this study.

name	sequence 5`-3`	application
AS_SS_RT	GACTGGAGCACGAGGACACTGCTAA AATCAATGTGTTGCATGT	Strand specific primer for synthesis of antisense by reverse transcription. Primer with an adapter sequence at its 5' end (Fedak et al., 2016).
AS_F	GACTGGAGCACGAGGACACT	PCR forward primer for the amplification of antisense transcript. Sequence of the adapter on *asDOG1* RT primer (Fedak et al., 2016).
AS_R	ACGTTAGGCTCTCCGACATT	PCR reverse primer for the amplification of antisense transcript (Fedak et al., 2016).
*UBC*1	CTGCGACTCAGGGAATCTTCTAA	PCR forward primer for the refence gene *UBC* (Czechowski et al., 2005).
*UBC*2	TTGTGCCATTGAATTGAACCC	PCR reverse primer for the refence gene *UBC* (Czechowski et al., 2005).
*PP2A*_F	TATCGGATGACGATTCTTCGTGCAG	DNase treatment efficiency check. PCR forward primer for *PP2A* (Fedak et al., 2016).
*PP2A*_R	GCTTGGTCGACTATCGGAATGAGAG	DNase treatment efficiency check. PCR reverse primer for *PP2A* (Fedak et al., 2016).
*DOG1*_total_F	AGCTCAACGACGATCTCAC	PCR forward primer for the amplification of *DOG1* exon1 (Fedak et al., 2016).
*DOG1*_total_R	ACATCGGTGAGCAAGATCAG	PCR reverse primer for the amplification of *DOG1* exon1 (Fedak et al., 2016).
*AtGP1*_F	TGGTTTTTCCTGTCCAGTTTG	PCR forward primer for the amplification of *GP1* (CPT positive control) (Dinh et al., 2014).
*AtGP1*_R	AACAATCCTAACCGGGTTCC	PCR reverse primer for the amplification of *GP1* (CPT positive control) (Dinh et al., 2014).
DOG1 Promoter (1)_F	TGGAACAACAACTCGCACTC	DRIP-qPCR forward primer for *DOG1* promoter region.
DOG1 Promoter (1)_R	CCGAGGAAATAAAAGAAATAACG	DRIP-qPCR reverse primer for *DOG1* promoter region.
DOG1 Promoter (2)_F	TTTGTGAGTGTGTCGGCTTC	DRIP-qPCR forward primer for *DOG1* promoter region.
DOG1 Promoter (2)_R	GAGAGTGCGAGTTGTTGTTCC	DRIP-qPCR reverse primer for *DOG1* promoter region.
DOG1 exon 1 (1)_F	TTCCACGTGGGTGCATAATA	DRIP-qPCR forward primer for *DOG1* exon 1.
DOG1 exon 1 (1)_R	GCTCAACGACGATCTCACG	DRIP-qPCR reverse primer for *DOG1* exon 1.
DOG1 exon 1 (2)_F	AGCTCAACGACGATCTCAC	DRIP-qPCR forward primer for *DOG1* exon 1.
DOG1 exon 1 (2)_R	ACATCGGTGAGCAAGATCAG	DRIP-qPCR reverse primer for *DOG1* exon 1.
DOG1 exon 1 (3)_F	GAGCGTTCTCTAAAGGACTGTTCCAC	DRIP-qPCR forward primer for *DOG1* exon 1.
DOG1 exon 1 (3)_R	GAGCTCAAACAACTCTTAGCTCAACG	DRIP-qPCR reverse primer for *DOG1* exon 1.
DOG1 intron 1 (1)_F	AGTACGGTGCGGCAAAAA	DRIP-qPCR forward primer for *DOG1* intron 1.
DOG1 intron 1 (1)_R	TTCCAAATTCAAACCGAACC	DRIP-qPCR reverse primer for *DOG1* intron 1.
DOG1 intron 1 (2)_F	AGGGTTTGGACGTTTTCGGTT	DRIP-qPCR forward primer for *DOG1* intron 1.

DOG1 intron 1 (2)_R	CCGTACTGACTACCGAACCA	DRIP-qPCR reverse primer for *DOG1* intron 1.
DOG1 ex2-int2 (1)_F	TGCATGAGTGGGGAACTATG	DRIP-qPCR forward primer for *DOG1* exon 2-intron 2 junction.
DOG1 ex2-int2 (1)_R	TTATGCAATTTTAAATATGACACGTA	DRIP-qPCR reverse primer for *DOG1* exon 2-intron 2 junction.
DOG1 ex2-int2 (2)_F	AACGACTACTTTCCTTCCTCTCC	DRIP-qPCR forward primer for *DOG1* exon 2-intron 2 junction.
DOG1 ex2-int2 (2)_R	TCGTGACTGTATGGTTGACACC	DRIP-qPCR reverse primer for *DOG1* exon 2-intron 2 junction.
DOG1 intron 2_F	CTGTATTTCGCAAAATGCCACGACGT	DRIP-qPCR forward primer for *DOG1* intron 2.
DOG1 intron 2_R	GTTTCGTTATAAGATTGTAGTTTGTAAGGA	DRIP-qPCR reverse primer for *DOG1* intron 2.
DOG1 int2-ex3_F	TCGAGACGAGATCATGTTGC	DRIP-qPCR forward primer for *DOG1* intron 2-exon 3 junction.
DOG1 int2-ex3_R	TCACGTCGTGGCATTTTG	DRIP-qPCR reverse primer for *DOG1* intron 2-exon 3 junction.
DOG1 exon 3_F	CCCACGGAGACGACAAATAATG	DRIP-qPCR forward primer for *DOG1* exon 3.
DOG1 exon 3_R	TTGTCGAGACGAGATCATGTTG	DRIP-qPCR reverse primer for *DOG1* exon 3.
DOG1 3' UTR_F	CGTCTCGACAAGTCAGCTAGG	DRIP-qPCR forward primer for *DOG1* 3' UTR.
DOG1 3' UTR_R	AAAAAGGATGCTTCCAACAA	DRIP-qPCR reverse primer for *DOG1* 3' UTR.

Table S2 – Table of Ct values for the tested primer pairs throughout *DOG1*. qPCR with 10% input samples with primer pairs throughout *DOG1 locus*. Ct values represent the mean of obtained Ct values from three biological replicates. In bold are the primer pairs chosen for the following DRIP-qPCR experiments. * represent the primer pairs designed in this work.

Primer pairs for *DOG1* region	Ct values
DOG1 Promoter (1)*	22.98
DOG1* Promoter (2)	**21.57**
DOG1 exon1 (1)	22.93
***DOG1* exon1 (2)**	**22.13**
DOG1 exon1 (3)	25.44
Intron 1 (1)	24.12
Intron 1 (2)*	**21.88**
DOG1 ex2-int2 junction (1)	24.08
DOG1* ex2-int2 junction (2)	**22.18**
DOG1 intron 2	28.77
DOG1* int2-ex3 junction	**21.73**
***DOG1* exon 3**	**21.99**
DOG1 3' UTR	25.81

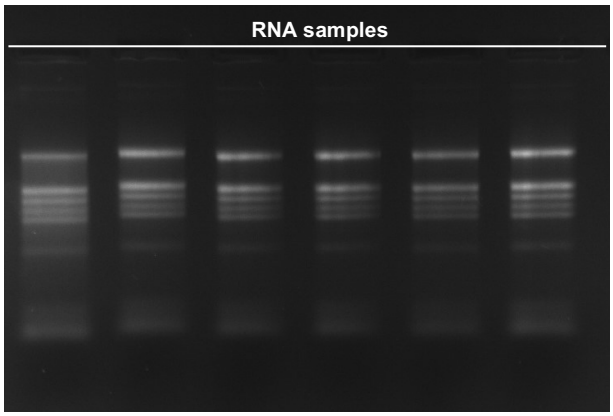


Fig. S1 - **RNA samples considered to be of good quality**. Representative agarose gel picture of RNA samples without signs of degradation and strong genomic DNA contamination. Samples from 5 days-old Col-0 seedlings treated with DMSO (control; three biological replicates; first three lanes starting from the left to the right) and 25 µM of CPT for 36h (last three lanes at the right). 100 ng of each RNA sample was loaded on the gel.

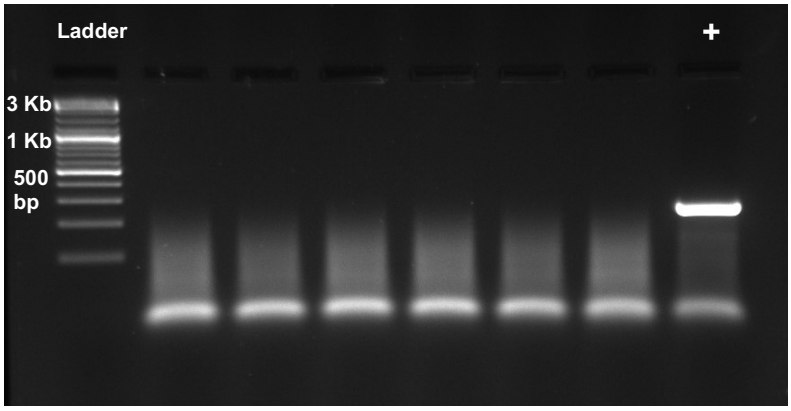


Fig. S2 – **DNA digestion confirmation**. Agarose gel image after PCR with primers for *PP2A* gene (*At1G69960.1*) on the RNA samples from 5 days-old Col-0 seedlings treated with DMSO and 25 µM of CPT for 36h (samples from Fig. S1) treated with DNase I. DNA ladder on the first lane, 6 samples run on lanes 2 to 7, and one positive control (+; RNA sample used on PCR not treated with DNase I) showing amplification of the genomic *PP2A* DNA sequence on the last lane, from left to right.

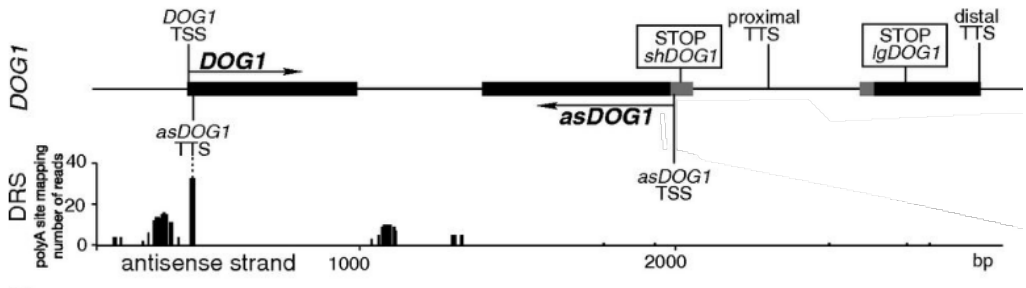


Fig. S3 – **Reanalysis of polyA site mapping by Direct RNA sequencing** (Sherstnev et al., 2012). Reads mapped to the antisense strand represent sites where polyadenylation occurs (*asDOG1* TTS) adapted from (Fedak et al., 2016).

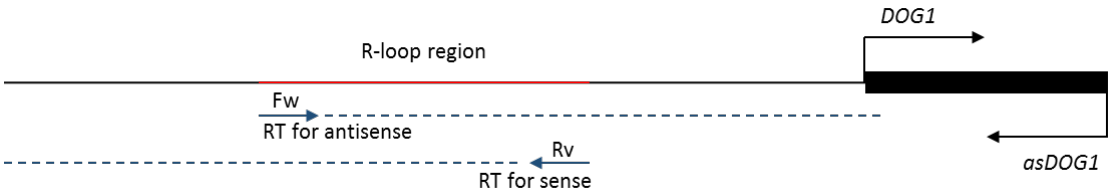


Fig. S4 – **Strategy to identify the orientation of the RNA forming the R-loop at *DOG1* promoter**. Reverse transcription with the forward primer (Fw) leads to the synthesis of cDNA from an antisense transcript. Reverse transcription using the reverse primer (Rv) leads to the synthesis of cDNA from a sense transcript. PCR with both primers on the Fw or Rv cDNA samples reveals the orientation of the RNA at the R-loop region.