Charles University

Faculty of Science

Department of Experimental Plant Biology

PŘÍRODOVĚDECKÁ FAKULTA
Univerzita Karlova

Vojtěch Čermák

Dynamics and variability of induced transgene silencing in tobacco cell line BY-2

Dynamika a variabilita indukovaného umlčování transgenů v tabákové buněčné linii BY-2

Doctoral thesis

Supervisor: Lukáš Fischer

Prague, 2021

DECLARATIONS

I hereby declare that this Ph.D. thesis documents my own work and I wrote it independently. This work or any substantial part of the text is not a subject of any other defending procedure. I declare that all used sources were cited and acknowledged properly.

Prague, 30. 7. 2021 .....
Vojtěch Čermák

In the name of other co-authors, I declare that Vojtěch Čermák has substantially contributed to all selected publications and I agree with the fact that these articles are presented as an integral part of this Ph.D. thesis. The publications were created by collectives of authors and the participation of the author of this thesis is specified below.

Prague, 30. 7. 2021 .....
Lukáš Fischer

PROHLÁŠENÍ

Prohlašuji, že tato disertační práce představuje mou vlastní práci a psal jsem ji samostatně. Tato práce nebo její jakákoli podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu. Dále prohlašuji, že všechny použité zdroje byly správně citovány.

V Praze, 30. 7. 2021 .....
Vojtěch Čermák

Ve jménu dalších spoluautorů prohlašuji, že Vojtěch Čermák významně přispěl do všech vybraných publikací, a souhlasím s tím, že tyto články jsou prezentovány jako nedílná součást této disertační práce. Publikace byly vytvořeny kolektivy autorů a účast autora této práce je uvedena níže.

V Praze, 30. 7. 2021 .....
Lukáš Fischer

Acknowledgments

I would like to thank my supervisor Lukáš Fischer for his guidance, fruitful discussions and friendly leadership. I would also like to thank all the members of the Department of Experimental Plant Biology, especially the members from the laboratory of Plant Cell Biology and Biotechnology for all the help they provided and the friendly environment. Special thank belongs to the people who were in any way involved in projects presented in this thesis, these are namely: Míťa Tyč, Adéla Přibylová, Šárka Motylová, Lenka Dvořáková and Eliška Kobercová.

Finally I would like to thank my family and my wife for their support.

Financial support

Experiments presented in this work were done at the Department of Experimental Plant Biology, Faculty of Science, Charles University. The work was supported by the Ministry of Education, Youth and Sports of Czech Republic (project IDs: LO1417 and LM2015042) and by Charles University (GAUK 904813).

Content

Abstract

RNA interference (RNAi) is an important mechanism regulating gene expression. In plants, RNAi is triggered by double-stranded RNA (dsRNA) which is processed into small RNAs (sRNAs), usually 21-24 nt long. The sRNAs are loaded into Argonaut (AGO) protein and recognize the target based on sequence complementarity. When the target is mRNA, they can slice it or block translation leading to posttranscriptional gene silencing (PTGS). When the target is DNA, they can induce DNA methylation and chromatin changes, which when present in the promoter can lead to transcriptional gene silencing (TGS). The individual components of RNAi are well described, but less is known about the impact of different types of dsRNA precursors on the dynamics of RNAi.

To study these aspects of RNAi, we used tobacco BY-2 cell line expressing GFP reporter and inducible silencers. The silencers used different ways of triggering the dsRNA formation by transcripts from antisense (*AS*), unterminated sense (*UT*) and inverted repeat (*IR*) *GFP* sequence to initiate PTGS. Additionally, one *IR* silencer based on the *CaMV 35S* promoter initiated TGS. This allowed us to study RNAi from the beginning throughout the steady state level and till the recovery phase, all in the highly homogeneous system.

Using this system, we described several features of silencing with also some unexpected observations. First of all, we show that in some cases the silencing can be initiated by transgenes without any promoters, just by read-through transcription from outside the *T-DNA*. Then, we provide a detailed description on the dynamics of PTGS triggered by the three silencers with different mechanisms of forming dsRNA. We show that the dynamics and the effects of the silencing differ between these silencers. The spectrum of produced sRNAs also differs between silencers, but not enough to explain all the differences in silencing. Our results call into question the mechanism of silencing triggered by aberrant RNAs (*UT* silencer): the nature of these transcripts and the conditions under which they are converted to the dsRNA. We observed that *de novo* methylation triggered by sRNAs can differ between two loci with the same sequence, likely as a result of different transcription history of these loci. We also describe the initiation of TGS in detail. We show that DNA methylation and the resulting TGS can be quite fast and during this TGS, we observed different dynamics of methylation for the particular cytosine contexts. Our results provide a comprehensive description of the dynamics of silencing in plants and raise new questions about several aspects of plant RNAi.

Abstrakt

RNA interference (RNAi) je důležitým mechanismem regulace genové exprese u rostlin. RNAi u rostlin je spouštěna přítomností dvouvláknové RNA (dsRNA), která je následně zpracována na malé RNA (sRNA) o délce 21-24 nt. Malé RNA po navázání na protein Argonaut (AGO) jsou schopné rozeznat cílovou molekulu na základě sekvenční komplementarity. Pokud je cílem mRNA, může ji AGO rozštěpit, nebo může blokovat translaci, což vede k posttranskripčnímu umlčení (PTGS). Pokud je cílem DNA, tak může navodit její metylaci, což vede k transkripčnímu umlčení (TGS), pokud dojde k metylaci v promotoru. Jednotlivé komponenty RNAi jsou celkem dobře popsány, ale méně je známo o vlivu odlišných typů dsRNA prekursorů na dynamiku RNAi.

Ke studiu těchto aspektů RNAi jsme využili tabákovou buněčnou linii BY-2 s reportérovým genem pro GFP cíleným silencery s inducibilní expresí. Fungování silencerů bylo založeno na třech základních způsobech tvorby dsRNA: transkripcí antisense *GFP* (*AS*), *GFP* bez terminátoru (*UT*) a *GFP* v invertované repetici (*IR*). Tyto tři silencery spouštěly PTGS, čtvrtý silencer byla *IR* cílící na *CaMV 35S* promotor a spouštějící TGS. To nám umožnilo studovat RNAi od její iniciace, přes plné umlčení až po obnovení exprese při vysazení induktoru – vše relativně snadno analyzovatelné díky vysoce homogennímu systému.

S použitím tohoto systému jsme detailně popsali průběh umlčování s některými jeho neočekávanými vlastnostmi. Nejprve jsme ukázali, že je možné za určitých okolností spustit umlčování i z transgenu bez promotorů, jen díky pročítání transkriptů z vnějšku T-DNA. Dále jsme detailně zmapovali dynamiku PTGS spouštěného třemi různými silencery s různými způsoby formování dsRNA. Dynamika a důsledky umlčování se mezi všemi silencery lišily. Rozdíly byly i v množství a spektru produkovaných sRNA, ale tyto rozdíly nebyly dostačující k vysvětlení všech rozdílů v průběhu umlčování. Z našich výsledků také vyvstávají otázky ohledně umlčování spouštěného aberantními RNA (*UT* silencer) – jaká je povaha těchto transkriptů a za jakých podmínek z nich dsRNA vznikají? Pozorovali jsme rozdílnou citlivost dvou lokusů se stejnou sekvencí k *de novo* metylaci pomocí sRNA, pravděpodobně v důsledku odlišné historie transkripce na těchto lokusech. V neposlední řadě jsme také detailně popsali počáteční fáze TGS. Toto TGS a metylace s ním spojená byly indukovány velmi rychle, dynamika metylace se ale s ohledem na kontexty jednotlivých cytosinů lišila. Námi získaná data přináší velmi detailní pohled na dynamiku umlčování u rostlin a nastolují některé nové otázky o RNAi u rostlin.

Abbreviations

AGO	Argonaut
DCL	Dicer-like
dsRNA	double-stranded RNA
GFP	Green fluorescent protein
miRNA	micro RNA
NAT	natural antisense transcript
Pol	DNA-dependent RNA polymerase
PTGS	posttranscriptional gene silencing
RdDM	RNA-directed DNA methylation
RDR	RNA-dependent RNA polymerase
RNAi	RNA interference
siRNA	small interfering RNA
sRNA	small RNA
TGS	transcriptional gene silencing

Introduction

Genes are essential building blocks of living organisms. For genes to function meaningfully, it is crucial that their activity is properly regulated. Gene activity in eukaryotic cell can be controlled at many levels, beginning with transcriptional regulation (i.e. whether the transcription alone gets initiated) and continuing through all the steps to the final protein, which itself can be further modified or degraded.

One of the mechanisms that is able to regulate gene expression is RNA interference (RNAi). The key element of RNAi are small RNAs (sRNAs) along with their effector Argonaut (AGO) proteins (Fig. 1). These sRNAs allow RNAi to define its target based on sequence complementarity, making the RNAi pathway a versatile mechanism that can regulate various processes where the target is nucleic acid. Indeed the RNAi pathway has been described in regulating processes such as gene expression, modifying chromatin, directing defense against viruses and transposons and DNA damage repair to name a few (Borges & Martienssen 2015). The most prominent and studied functions of RNAi are regulation of gene expression

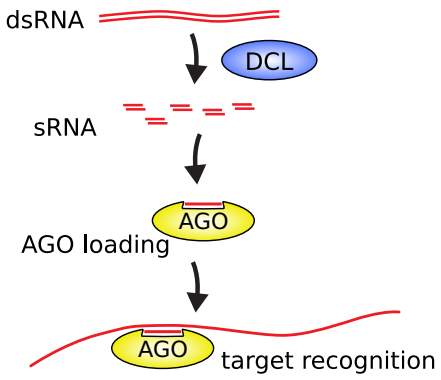


Fig. 1: The basic mechanism of RNAi in plants.

and defense against invading nucleic acids (viruses, transposons and transgenes).

In plants the RNAi pathway is quite diverse. It has all the key enzymes of eukaryotic RNAi, but it also possesses several features that developed specifically in plants and which are mostly involved in regulation of chromatin (Lee & Carroll 2018). Plants are also important as model organisms to study RNAi, they were essential to the discovery of RNAi (Matzke & Matzke 2004) and to this date they are still leading the way of RNAi research.

Small RNA biogenesis

In plants sRNAs are generated from double-stranded RNA (dsRNA) precursors. The dsRNA as a trigger for RNAi is common to all eukaryotes with only few exceptions, where sRNAs can be produced from single-stranded RNA (ssRNA) precursor, however, it seems that such pathway does not exist in plants (Law & Jacobsen 2010).

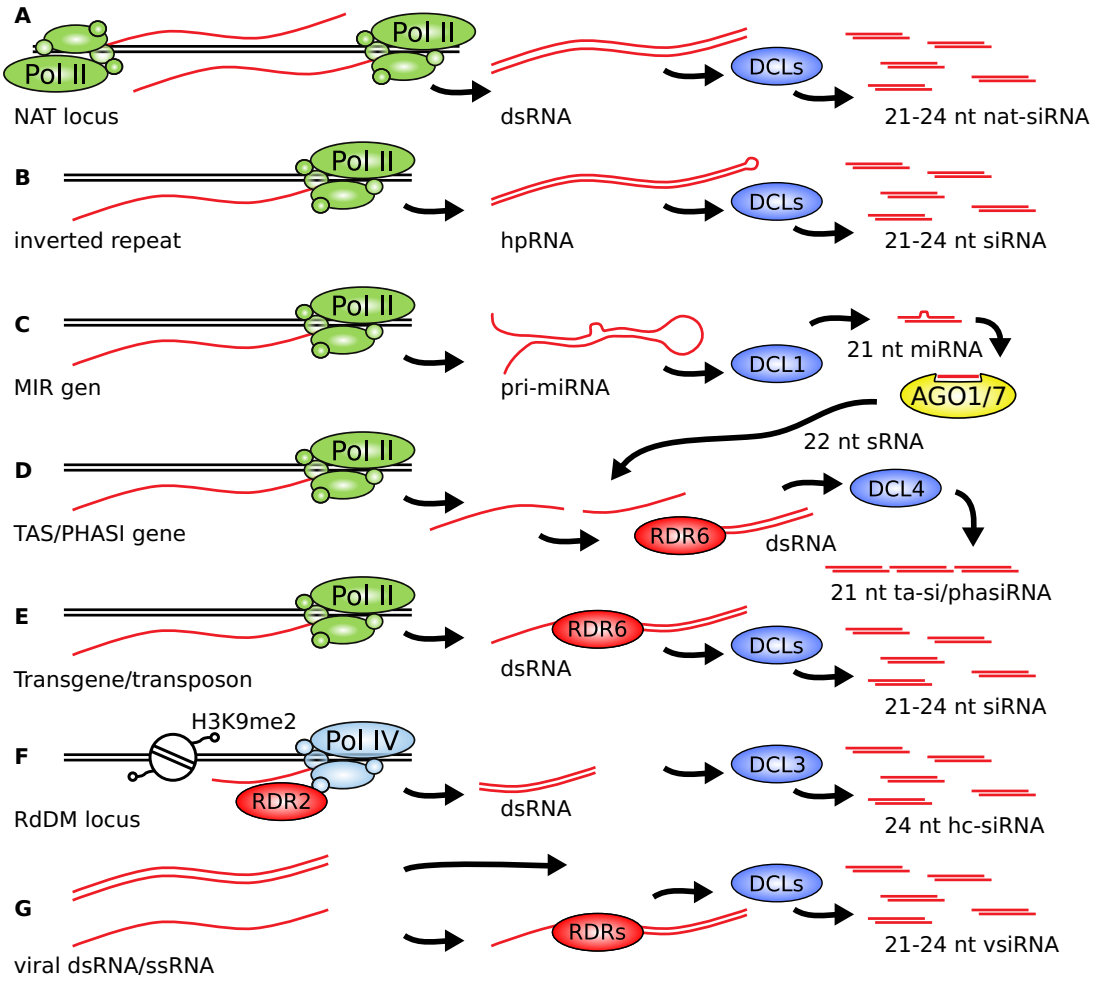


Fig. 2: Schematic representation of various pathways of sRNA biogenesis in plants. The most common gene paralogs for given pathway are indicated. A) Production nat-siRNA, the complementary transcript can be also produced in trans; B) siRNA derived from long inverted repeats; C) miRNA pathway; D) phasi/ta-siRNA pathway – usually triggered by 22nt miRNA/siRNA, generally all secondary siRNA are produced in similar fashion; E) RNAi triggered by aberrant RNAs; F) Pol IV pathway of heterochromatic-siRNA production; G) production of viral siRNA.

There are many ways the dsRNA precursor can be formed. Generally the primary ssRNA is produced by DNA-dependent RNA polymerase (usually Pol II or plant specific Pol IV) or it can be of foreign origin (e.g. viral RNA). This RNA then has to form dsRNA, which can happen through intramolecular paring (by forming hairpins if inverted repeat sequences are present within the molecule), intermolecular pairing (by pairing of two complementary transcripts) or by activity of RNA-dependent RNA polymerase (which can synthesize the second strand to a single-stranded RNA). The precursor, its biogenesis and processing is fundamental for sRNAs classification (Fig. 2). Although only few sRNAs classes are defined precisely, while many others have more loose definitions probably given the multiple redundancies in the many pathways of their biogenesis.

Precursors for small RNAs

Probably the most straightforward way of producing dsRNA is transcription of inverted repeats. These when transcribed can fold back on itself to create a hairpin structure. There are several long inverted repeats in the *Arabidopsis* genome producing heterogeneous groups of small interfering RNA (siRNA; Zhang et al. 2007; Polydore & Axtell 2018). Much more abundant, however, are shorter repeats which give rise to so called micro RNAs (miRNA). miRNAs make a well-defined class of sRNAs. Their precursors (pre-miRNA) have specific secondary structure based on a stem-loop with imperfect pairing (Axtell & Meyers 2018). They are transcribed by Pol II from MIR genes and one specific miRNA can be produced from several distinct genes (these are so called miRNA families). Many miRNAs are conserved throughout land plants and they are important posttranscriptional regulators of gene expression (Axtell & Meyers 2018). Given the specific features of miRNA pathway, sRNAs are often divided into two groups: miRNA and siRNA, however, these two categories are not equal and the siRNA group is heterogeneous and contains various sRNAs categories.

The alternative possibility to intramolecular pairing is intermolecular pairing. Many genes (either coding or non-coding) do have antisense transcripts (natural antisense transcripts – NATs), which are produced either *in cis* or *in trans* (more than 20% of *Arabidopsis* genes). Such two RNA molecules have the possibility to base pair and generate dsRNA. However, only few NATs seem to be capable of producing nat-siRNAs, the rest of them have other functions unrelated to RNAi serving often as scaffold RNAs (Zhang et al. 2012; Yuan et al. 2015; Wierzbicki et al. 2021). Most of the known nat-siRNAs are involved in gene regulation in response to stress (Zhang et al. 2012).

The third way which can lead to dsRNA formation is to synthesize the second strand based on the single-stranded RNA. Plant genomes code for several RNA-dependent RNA polymerases (RDR).

Of the three eukaryotic RDR clades, two are known in plants. While the function of RDRγ in plants remains unknown, the RDRα polymerases are an essential component of plant RNAi (Lee & Carroll 2018). In fact most plant sRNAs are dependent on the activity of RDRs (Polydore & Axtell 2018). In *Arabidopsis* there are three such polymerases: RDR1, RDR2 and RDR6. RDR1 is important in antiviral defense (Diaz-Pendon et al. 2007), RDR2 in transcriptional silencing and DNA methylation (Kasschau et al. 2007; Stroud et al. 2013; Mishra et al. 2021) and RDR6 in generating secondary siRNAs and posttranscriptional gene silencing (Yoshikawa et al. 2005; Zheng et al. 2010). While these RDRs are redundant when it comes to antiviral defense, for most other pathways they are quite specific (Kasschau et al. 2007). It is important that the RDRs target RNAs very specifically, otherwise many important genes could be accidentally silenced. There are several mechanisms that can target RDRs to specific RNAs.

RNAs can be produced specifically to become RDR substrate. Plants have evolved DNA-dependent RNA polymerase IV (Pol IV) whose transcripts are directly channeled to RDR2 to produce dsRNA precursors (Mishra et al. 2021). Along with Pol V (described later), Pol IV is plant specific polymerase which evolved from Pol II to be involved in RNA-directed DNA methylation (RdDM). Pol IV is targeted to chromatin with specific features (like H3K9me2 modification), where it produces short transcripts, each of them being a precursor for single hc-siRNA (heterochromatic-siRNAs or p4-siRNA; Law et al. 2013; Zhai et al. 2015; Zhou et al. 2018). As these siRNAs are key players in transposon silencing, they make up most sRNAs present in plant cells (Mosher et al. 2008).

Besides Pol IV transcripts, Pol II transcripts can also serve as RDR substrate. Pol II transcripts are mostly recognized by RDR6, however, the route to dsRNA is not as direct as with Pol IV transcripts. This phenomenon is best studied on transgenes, where RDR6 dependent RNAi is the major obstacle to stable transgene expression in plants. The current paradigm is that RDR6 substrate are aberrant transcripts that escape conventional RNA degradation pathways. Transgenes are often source of aberrant RNAs because of common flaws in the construct design, the major factor in this is inappropriate use of promoters and terminators, but other factors can also contribute, like the absence of splicing or suboptimal codon usage (Luo & Chen 2007; Christie et al. 2011; Sidorenko et al. 2017; Felippes et al. 2020). In this model the use of stronger promoter in combination with not so efficient terminator will lead to production of RNAs lacking the 3’ poly-A tail and in some cases also read-through uncapped transcripts overlapping downstream genes. These transcripts, when produced in large enough quantities so they are not quickly enough removed by exonucleases, are recognized by RDR6 and enter the RNAi pathway (Herr et al. 2006; Luo & Chen

2007; Moreno et al. 2013; Parent et al. 2015b; Krzyszton et al. 2018). The important role of RNA degradation pathways can be seen in various mutants in these pathways, where many native genes become silenced resulting in severe phenotypes that can be mostly rescued by mutation in *rdr6* (Martínez de Alba et al. 2015). RDR6 is also directly inhibited by the presence of a poly-A tail (Baeg et al. 2017). Although this phenomenon is mostly studied on transgenes it has been recently suggested that suboptimal codon usage in transposable elements can lead to aberrant RNAs production and thus subject transposon transcripts to RNAi (Kim et al. 2021). Therefore recognizing aberrant RNAs by RNAi may be a general mechanism how plants can protect itself from invasive genic elements.

Besides direct production of aberrant RNAs by Pol II, it is also possible to generate aberrant RNAs posttranscriptionally by sRNA directed cleavage of mRNA. The sRNA-targeting can thus lead to production of secondary siRNAs. These secondary siRNAs are generated from the 5' and 3' cleavage fragments that lack the mRNA cap and poly-A tail, respectively, and that can be converted to dsRNA by RDR6. However, not all sRNA targets produce secondary siRNAs and those that do differ in the quantity and quality of resulting secondary siRNAs (Zhang et al. 2015). It has been shown that the production of secondary siRNAs can be programmed by the primary sRNA – in case of miRNA by the structure of their precursor and/or length (22 nt) and in case of siRNA by their length (22 nt) and/or biogenesis (Mlotshwa et al. 2008; Manavella et al. 2012). The production of secondary siRNAs has many important implications for the function of RNAi in plants – it allows for the amplification of silencing signal (with positive feedback loop) and also to spread the production of siRNAs along the target transcript in process called transitivity, which can result in targeting cascade of new genes just by homology with the primary target. In some cases this amplification step seems to be essential for efficient silencing, as it appears that the nat-siRNA and probably even the antisense RNA technology are both dependent on the activity of RDRs (Parent et al. 2015b; Yuan et al. 2015; Polydore & Axtell 2018).

Processing of small RNAs

After the dsRNA is generated as described above, it can be processed into sRNAs. This is done by RNase III enzyme Dicer, in plants called Dicer-like (DCL). For sRNA processing, three DCL domains are the most important: two RNase III domains and a PAZ domain. The RNase III domains cleave the dsRNA in a way so they leave 2 nt 3' overhang. The PAZ domain recognizes this 2 nt 3' overhang and positions the DCL so it can make a cut in a distance which is equal to the distance between the RNase III domains and PAZ domain for a given DCL enzyme. This releases dsRNA (sRNA duplex) of specific length with 2 nt 3' overhang at both ends. One of the strands from this

duplex will later become the functional single-stranded sRNA (described below). The DCL proteins are assisted by DRB (double-stranded RNA-binding) proteins in this process. There are usually four *DCLs* genes in dicot plants and five in monocot plants, each with preference for different dsRNA substrate and producing sRNAs of specific length (Hiraguri et al. 2005; MacRae et al. 2006; Margis et al. 2006).

DCL1 generates 21 nt long sRNAs and it is primarily responsible for miRNA processing where it shows little or no redundancy with other DCLs. Beside miRNAs it can also contribute to processing of other sRNAs like nat-siRNAs (Ronemus et al. 2006; Yuan et al. 2015). DCL2 generates usually 22 nt long sRNAs which are processed from various long dsRNA precursors. These 22 nt long sRNAs have the ability to induce production of secondary siRNAs. However, since DCL2 is subordinate to DCL4, this ability is usually suppressed. In *dcl4* mutant the DCL2 dependent secondary siRNAs will begin to target several endogenous genes that would not be normally silenced (Mlotshwa et al. 2008; Parent et al. 2015a; Wu et al. 2017). DCL3 usually generates 24 nt long sRNAs, most of which are processed from the dsRNAs produced by the Pol IV-RDR2 pathway and are involved in the process of RdDM. It can also contribute to the processing of other dsRNA substrates, but usually only in subordinate role to DCL4 and DCL2. The short Pol IV-RDR2 precursors can by also processed by other DCLs into 21-22 nt sRNAs and surprisingly also by DCL independent pathway using RNA exonucleases (Kasschau et al. 2007; Marí-Ordóñez et al. 2013; Ye et al. 2016; Panda et al. 2020). DCL4 generates 21 nt long sRNAs from long dsRNA precursors and in takes a precedence in their processing over other DCL proteins. It also has the ability to produce sRNAs in phase – meaning that the sRNAs are produced almost exclusively in 21 nt intervals from the start point of processing. These are called phasiRNA, some of which make ta-siRNA (*trans*-acting siRNA) subgroup. The ta-siRNAs are produced from specialized non-coding RNAs and their processing requires previous slicing by other sRNA, usually miRNA. The phasiRNA are important in morphogenesis and gametogenesis where they also exist in 24 nt form processed by DCL3 or its paralogs in some species (Yoshikawa et al. 2005; Fei et al. 2013; Nagano et al. 2014; Parent et al. 2015a; Arribas-Hernández et al. 2016; Xia et al. 2019). When the final sRNA duplex is diced out of the precursor it is 2'-O-methylated at the 3' ends by HEN1 as a protection against degradation (Li et al. 2005).

All the RNAi components discussed so far localize to the nucleus, predominantly to the nucleolus-associated Cajal bodies. Only some of these proteins also have cytoplasmic localization (RDR6, DCL2, DCL4…). This is interesting because many of the substrate RNAs are present in cytoplasm. Thus the compartmentalization of RNAi processes still needs many answers, but the

same spatial localization of RNAi components is likely connected with the high redundancy especially among the DCL proteins (Hoffer et al. 2011; Pontes et al. 2013; Pumplin et al. 2016).

Roles of small RNAs in silencing

After the sRNA is produced by one of the pathways described above, it can then serve its purpose. But it cannot act alone. The second key component in the RNAi pathway along with the sRNA is protein Argonaut (AGO). The AGO protein loads the sRNA to become RNA Induced Silencing Complex (RISC) and this ribonucleoprotein is the effector of RNAi. Not all sRNAs can be loaded into all AGOs. The AGO selects its specific sRNA based on the length of the sRNA duplex, identity of the 5' nucleotide and thermodynamic stability at the 5' end (Mi et al. 2008; Takeda et al. 2008; Eamens et al. 2009; Havecker et al. 2010). Only one of the strands (guide strand) from the duplex is selected, while the other (passenger strand) is degraded – usually by the slicer activity of the AGO itself (Arribas-Hernández et al. 2016). The loaded sRNA allows AGO to recognize target RNA based on sequence complementarity. The AGO itself usually has RNase H-like function and is able to slice the target RNA molecule or it can mediate targeting of other enzymes. Besides general protein-protein interactions, AGOs also bind proteins containing an AGO-hook – specific disordered domain with WG/GW motives (El-Shami et al. 2007; Karlowski et al. 2010).

AGO is an evolutionary old protein already present in prokaryotes where it is a part of RNA-guided restriction systems. It adopted its function in RNAi already in the last eukaryotic common ancestor (along with other RNAi proteins like Dicer and RDR). In plants the AGO family is quite large, ranging from 10 genes in *Arabidopsis* to 27 in *Brassica napus* (Lee & Carroll 2018). Different AGOs load different (but sometimes overlapping) pools of sRNAs and they have specific functions in wide variety of processes. Their role in regulating gene expression can be separated into two levels: posttranscriptional and transcriptional gene silencing (PTGS and TGS, respectively).

Posttranscriptional gene silencing

As the name of posttranscriptional gene silencing (PTGS) suggests, this process acts at RNA transcripts (usually mature mRNA). It is mostly guided by 21-22 nt long sRNA (miRNA, nat-siRNA, ta-siRNA…) that can be produced by DCL1, 2 and 4. AGO1 is the primary AGO involved in PTGS and it also loads most miRNAs. AGO1 prefers sRNAs with 5' U. It is notable that AGO1 was the first discovered AGO and the *ago1* mutant phenotype of *Arabidopsis thaliana* give name to this protein family (Bohmert et al. 1998; Baumberger & Baulcombe 2005; Mi et al. 2008). AGO7 and AGO10 are involved in some specific miRNA regulated morphogenetic processes (Montgomery et al. 2008; Zhu et al. 2011). AGO2 and 5 associate with several miRNA, antiviral

siRNAs and have some other less defined functions. They prefer 5’ A and C respectively. AGO2 shows some redundancy with AGO1 (Takeda et al. 2008; Zhang et al. 2016).

As mentioned earlier, the AGO protein itself has slicer activity. This is sufficient for PTGS, the sliced mRNA can no longer be translated and the slicing products are subsequently degraded (Baumberger & Baulcombe 2005). Although most genes regulated by PTGS are regulated in this way, the mRNAs targeted by AGO protein can also have a different fate (Reis et al. 2015). At a subset of PTGS targets, the gene expression is regulated by blocking translation without slicing the mRNA. The AGO1 has been shown to associate with polysomes at endoplasmic reticulum. The exact mechanism deciding whether the target will be sliced or if the translation will be blocked is not known, but it is probably affected by the biogenesis and size of sRNAs. Several miRNAs (22 nt long and/or DRB2 dependent) and DCL2 dependent siRNA (22 nt long) have been shown to prefer translational inhibition which is involved mostly in regulating plant stress responses (Brodersen et al. 2008; Reis et al. 2015; Li et al. 2016; Wu et al. 2020).

Besides direct regulation of gene expression the PTGS can be also involved in secondary sRNAs production. As discussed before, the target RNA can become a substrate for the activity of RDR6 and this process is dependent on the origin of the primary sRNA. However, the AGO plays quite an active role in this process. AGO1 has been shown to recruit SGS3 protein which protects the sliced transcript from degradation and subsequently interacts with RDR6. This allows the secondary siRNA production machinery to overcome the RNA degradation pathways that would otherwise remove the sliced RNAs. It also turns out that in some cases the slicing is not even required for the secondary siRNA production (Zhang et al. 2015; Arribas-Hernández et al. 2016; Yoshikawa et al. 2016). This along with the similarities between sRNAs able to induce translational inhibition and secondary siRNA production (length of 22 nt) points to closer connection between translation and secondary siRNAs. Indeed, many substrates for secondary siRNA have been found to associate with polysomes, including supposedly noncoding precursors for ta-siRNAs – which as it turns out contain short open reading frames (ORFs) important for ta-siRNA production (Li et al. 2016; Yoshikawa et al. 2016).

Transcriptional gene silencing

Besides targeting already produced transcripts, the AGO-sRNA complex can be also targeted to chromatin and block transcription. At chromatin they can induce repressive modifications and if such modifications occur in the regulatory region of a gene (typically a promoter), such gene can get transcriptionally silenced. In plants the primary modification induced by RNAi is DNA

methylation (C5 methylation of cytosine) in a process called RNA-directed DNA methylation (RdDM; Fig. 3). This is usually guided by hc-siRNAs which are 24 nt long and produced by Pol IV-RDR2-DCL3 pathway. The DCL3-dependent 24 nt siRNA can be also produced from other types of precursors, for example when the preferred DCLs for their processing become saturated (Marí-Ordóñez et al. 2013). Although the 24 nt siRNAs are major effectors in RdDM, the 21-22 nt siRNAs can be also used in so called non-canonical RdDM (Pontier et al. 2012; Wu et al. 2012; McCue et al. 2015). AGO4 is the primary AGO involved in RdDM. It has three closely related *AGO* genes in *Arabidopsis*, *AGO6* and *9* are also involved in RdDM, but they show organ specific expression, while *AGO8* is considered to be pseudogene. All these AGOs prefer 24 nt siRNA with 5' A (Havecker et al. 2010). Although these proteins are very similar they show only partial redundancy, probably because of their expression patterns. Surprisingly AGO4 loads a very similar poll of sRNAs as AGO3 – a close paralog to AGO2 involved in PTGS. Both these AGOs prefer 5' A, but AGO2 binds 21 nt sRNAs, while AGO3 24 nt sRNAs (Havecker et al. 2010; Duan et al. 2015; Zhang et al. 2016). Although it seems that AGO2 is also involved in methylating DNA at small subset of loci (Pontier et al. 2012).

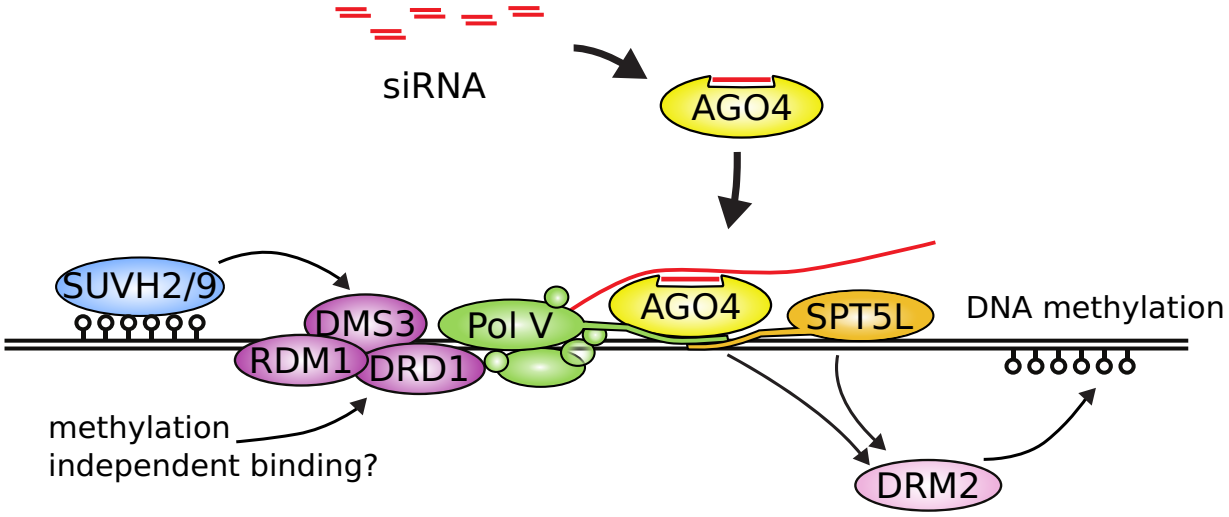


Fig. 3: RNA-directed DNA methylation (RdDM): Pol V is targeted to chromatin by DNA methylation and possibly also by other means. Access to chromatin is facilitated by chromatin remodeling complex (DRD1, DMS3 and RDM1). Pol V scaffold transcript is recognized by AGO bound siRNA. The AGO also interacts with Pol V and its transcription factor SPT5L. This results in recruitment of DRM2 and DNA methylation.

In PTGS the sRNA recognition of single-stranded target RNA is direct. In TGS the recognition of target DNA is not so straightforward. Plants solved this by specialized DNA-dependent RNA polymerase V (Pol V). This polymerase produces scaffold transcripts which can be easily targeted by sRNAs and in this way they can bring AGO to chromatin. Pol V is related to Pol IV (mentioned earlier) and both are derived evolutionary from Pol II. In non-plant organisms TGS is facilitated by

Pol II, but for some reason, plants separated this pathway from Pol II and are able to produce both sRNAs and scaffold transcripts without Pol II (Luo & Hall 2007; Bhattacharjee et al. 2019). Still, all these polymerases are closely related and share many subunits, the main difference is in the largest and second largest subunits, designated NRPD1 (Pol IV), NRPE1 (Pol V) and NRPD/E2 (the second largest subunit is shared by Pol IV and V in *Arabidopsis*). These subunits diverged from Pol II mainly in the catalytic center, and in the C-terminal domain (CTD) of the largest subunit. The CTD lacks the classical heptad repeats of Pol II and instead it has a DeCL domain of important but unknown function, which is present in both polymerases. Pol V CTD also has an AGO-hook domain important in binding AGO with sRNA (El-Shami et al. 2007; Marasco et al. 2017; Wendte et al. 2017; Wendte et al. 2019a). In the current model of RdDM, Pol V generates scaffold transcripts (50-200 nt long) at the target locus. AGO binds these transcripts based on sequence complementarity to its sRNA and it also binds to Pol V CTD and SPT5L CTD (Pol V elongation factor that also contains an AGO-hook domain). The AGO then slices the Pol V transcript, but the reason for it is not known. As a result of this process, DRM2 (Domains Rearranged Methyltransferase) methylates cytosines in DNA in close proximity (Wierzbicki et al. 2008; Zhong et al. 2014; Böhmdorfer et al. 2016; Liu et al. 2018; Gallego-Bartolomé et al. 2019). For loci to be methylated in this model, one needs an appropriate sRNAs but also Pol V transcription. Like in the case of Pol IV, it appears that Pol V does not need any specific promoter, but it is targeted via chromatin modifications. There is a specific set of proteins recognizing DNA methylation (SUVH2 and 9) which recruit chromatin remodeling complex (DRD1, DMS3 and RDM1) which in turn allows Pol V to start transcription (Johnson et al. 2014; Böhmdorfer et al. 2016; Gallego-Bartolomé et al. 2019). This raises a question of how loci with active chromatin marks can get *de novo* methylated? It has been recently suggested, that Pol V is able to some extent transcribe regions of the whole genome and thus allow sRNAs to target any loci for DNA methylation (Tsuzuki et al. 2020). This notion is supported by SUVH2 and 9 independent methylation of newly integrated DNA, however, more research on this subject is needed (Fultz & Slotkin 2017).

The key methyltransferase in RdDM is DRM2, it belongs to a family of plant specific methyltransferases structurally similar to mammalian Dnmt3. DRMs are able to methylate any cytosine and they are essential for the function of RdDM (Cao & Jacobsen 2002; Zhong et al. 2014). Besides DRMs plants have two other types of DNA methyltransferases. MET1 (Methyltransferasease1) maintains methylation in the symmetrical CG context after DNA replication based on the methylation status of the parental strand. MET1 is orthologous to mammalian Dnmt1. This methylation is important for heterochromatin maintenance and

methylation of gene bodies – methylation not related to silencing with unclear function, yet common to eukaryotes (Kankel et al. 2003; Cokus et al. 2008; Zemach et al. 2010). CMTs (Chromomethylases) maintain DNA methylation based on histone methylations, specifically H3K9me2. CMT2 methylates CHH sequences (H = A, C or T), but has a preference for CWA (W = A or T). CMT3 methylates symmetrical CHG with preference for CWG. They are important for heterochromatin maintenance (Zemach et al. 2013; Gouil & Baulcombe 2016; Stoddard et al. 2018). Although DRM2 is essential for RdDM, MET1 and CMT3 can also contribute by enhancing and maintaining methylation at RdDM targets in their respective sequence contexts (Cao et al. 2003; Aufsatz et al. 2004). So the primary role of RdDM is to establish DNA methylation and also to maintain it in euchromatin, while MET1 and CMTs function in maintaining already established methylation throughout the genome (Zemach et al. 2013; Zhang et al. 2018). Beside RdDM the ability of other pathways to establish *de novo* DNA methylation seems limited or nonexistent, however, recent studies have shown that CMTs might be also able to establish *de novo* DNA methylation in some cases; specifically in gene bodies, where the methylation is probably established at CHG sites by CMT3 to be later switched to CG and maintained by MET1 and also in mosses, where the RdDM is less prominent than in flowering plants (Yaari et al. 2019; Wendte et al. 2019b).

Besides establishment and maintenance, DNA methylation is also regulated by demethylation. The demethylation can be either passive (simply by replicating DNA without reestablishing the methylation at the daughter strand) or active (by enzymes). In plants the demethylation is done by glycosylases by removing the whole methylcytidine. The most prominent enzyme removing DNA methylation in somatic tissues is ROS1. The demethylation activity is likely guided by various chromatin modifications. These processes are important in plant reproduction and embryo development (Agius et al. 2006; Gehring et al. 2009; Tang et al. 2016).

The primary effect of DNA methylation on chromatin is induction of a repressive state. DNA methylation can directly affect binding of transcription factors to their DNA binding motifs. Over 70% of transcription factors are directly inhibited by DNA methylation, but there is small fraction of transcription factors with affinity to methylated DNA and some of these can actually activate expression when their binding site is methylated (Fischer et al. 2008; Lister et al. 2008; O'Malley et al. 2016; Harris et al. 2018). The best described example is the transcriptional regulation of ROS1. The ROS1 expression is positively correlated with levels of DNA methylation in its promoter, resulting in increased activity of ROS1 and DNA demethylation, creating a negative feedback loop. The decreased expression of ROS1 in RdDM mutants can lead to increased methylation at several

loci (Williams et al. 2015). However, DNA methylation usually does not act alone, but it is accompanied by other factors mostly in the form of chromatin modifications through which it can regulate the transcription. There are few best described chromatin modifications, which are directly regulated by DNA methylation and *vice versa*. The CHG methylation and to some extent also other methylated cytosines are recognized by SUVH4, 5 and 6 histone methyltransferases which induce H3K9me2 (dimethylation of histone 3 at lysine 9) modification. This modification is in turn recognized by CMTs to methylate DNA, creating a positive feedback loop, allowing CMTs to maintain DNA methylation. Pol IV also recognizes this modification, making the interconnection between DNA methylation and H3K9me2 very close (Bernatavichute et al. 2008; Law et al. 2013; Li et al. 2018). In opposition to H3K9me2 is H3K4me3, which is an activation mark. Several RdDM silenced loci require histone demethylase to remove this mark, so they can be silenced. H3K4me3 also blocks targeting of Pol IV (Greenberg et al. 2013; Law et al. 2013). Similar activation mark is histone acetylation and again RdDM is accompanied by histone deacetylases (Aufsatz et al. 2002; Liu et al. 2012). These marks directly affect the RdDM, but they act downstream of DNA methylation. In some other eukaryotes besides plants, the RNAi can also directly induce histone modifications (Cecere & Grishok 2014). Whether this is possible also in plants is not known, although few studies suggested the possibility of H3K9me2 being directly deposited by RNAi (Enke et al. 2011; Parent et al. 2021). There are many more chromatin modifications and histone variants which can have various effects on DNA methylation, transcription and other processes, making the chromatin regulation a very complex phenomenon (reviewed in Liu et al. 2010; Borg et al. 2021). These modifications are either recognized by transcription factors, or they can lead to chromatin compaction physically preventing transcription factors to bind DNA. The compaction is mostly the result of deacetylation and histone H1 binding. To methylate compacted DNA, DNA methyltransferases need the aid of chromatin remodelers (Zemach et al. 2013).

Aims of the Thesis

Silencing (RNAi) plays an important role in regulating gene expression and particularly it is of interest in case of transgene expression, because transgenes show high susceptibility to it. Although it has been intensively studied in recent years and many components of the RNAi pathway are known, there are still many aspects of silencing that lack adequate scientific description. One such area is the insight into the course of silencing, from its initiation to a steady state with the timing of individual steps. Studies on this topic analyzed silencing on the whole plant level, which makes the results closer to the real life conditions, but obscured by variability in tissues (each expressing different levels of RNAi components). In the long term experiments, there is also the impact of plant sexual reproduction, which brings large epigenetic resets with each generation (Paszkowski & Grossniklaus 2011; Zhao et al. 2014; Ingouff et al. 2017). To be able to approach these issues we chose a highly homogeneous tobacco BY-2 cell line as a model (Nagata et al. 1992) and established a system for monitoring the course of silencing and a system for precise silencing induction. We also decided to use different means of initiating RNAi, to compare how their dynamics of silencing differ.

The individual aims to achieve this were:

- *To establish a model system to monitor the course of silencing.* This would be a system that should allow us to monitor the silencing for theoretically unlimited time in unchanging and homogeneous culture. The silencing would have to be inducible to allow its monitoring from the very beginning and to switch it off. Such system also should be approachable by various methods to analyze it in sufficient detail.
- *To describe the time course of PTGS and TGS.* How fast is each type of silencing? Would they go through different stages with the respect to the strength of silencing, types of produced sRNA and the stability of silencing? Would they eventually reach maintenance phase, i.e. would PTGS eventually switch to TGS and would TGS eventually switch from RdDM to DNA methylation maintenance?
- *To find out differences in the course of PTGS and its outcomes between various silencers.* Assuming that different dsRNA precursors would lead to production of different types of sRNAs, in different amounts and with different spatiotemporal distribution, what would be their impact on PTGS? How does this affect speed and strength of PTGS, ability to methylate DNA and eventually switch to the maintenance phase of silencing?

Publication Summary

Publication 1 – Čermák & Fischer, 2018

Čermák, V. & Fischer, L., 2018. **Pervasive read-through transcription of T-DNAs is frequent in tobacco BY-2 cells and can effectively induce silencing.** BMC Plant Biology, 18(1), 252.

Summary: In this publication we describe silencing we observed in cell lines prepared for silencing induction with inverted repeat. Although the transcription of this inverted repeat should be inducible, we observed silencing often without induction. We show that there is a transcription independent of promoters present within the T-DNA and that it can be common for T-DNAs to be read-through by transcripts from genomic regions. This transcription when it encounters an inverted-repeat region can induce efficient silencing. We speculate that the read-through transcription is the result of a specific epigenetic status of the integrated T-DNA.

My contribution: I performed all the experiments and I significantly contributed to the study design, results interpretation and manuscript writing.

Publication 2 – Klíma et al. 2019

Klíma, P., **Čermák, V.**, Srba, M., Müller, K., Petrášek, J., Šonka, J., Fischer, L. & Opatrný, Z., 2019. **Plant Cell Lines in Cell Morphogenesis Research: From Phenotyping to -Omics.** In F. Cvrčková & V. Žárský, ed. Plant Cell Morphogenesis: Methods and Protocols. Methods in Molecular Biology. New York, NY: Springer New York, 367–376.

Summary: Methodological chapter about working with plant cell lines. One of the protocols describes flow cytometry analysis of BY-2 cell lines which was important in our works analyzing the course of silencing.

My contribution: I established the protocol and wrote description of the protocol “Assessing Fluorescence of BY-2 Protoplasts Using Flow Cytometry”

Publication 3 – Přibylová et al. 2019

Přibylová, A., **Čermák, V.**, Tyč, D. & Fischer, L., 2019. **Detailed insight into the dynamics of the initial phases of de novo RNA-directed DNA methylation in plant cells.** Epigenetics & Chromatin, 12(1), 54.

Summary: This work describes inducible TGS of *CaMV 35S* promoter in the tobacco BY-2 cell line and it is focused on its initial phases. We show that the methylation is very fast (first response

within 12 h), that different cytosine contexts have different dynamics of methylation and that preexisting CG methylation does not affect the time course of silencing.

My contribution: I was involved in the study design and some DNA methylation and transcript level analyses.

Publication 4 – Čermák et al. 2020

Čermák, V., Tyč, D., Přibylová, A. & Fischer, L., 2020. **Unexpected variations in posttranscriptional gene silencing induced by differentially produced dsRNAs in tobacco cells.** Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1863(11), 194647.

Summary: In this publication we compared different ways of inducing PTGS, specifically we compared: inverted repeat (IR), antisense (AS) and unterminated sense (UT) transcripts. All the silencers were inducible and targeted stably expressed GFP mRNA in tobacco BY-2 cell line. We observed that IR was fastest and strongest, UT was able to maintain silencing without ongoing induction and AS together with low induced IR had different impact on DNA methylation at two loci with identical sequence – indicating that these two loci had different sensitivity to RdDM.

My contribution: I did the majority of the experimental work and I significantly contributed to the study design, results interpretation and manuscript writing.

Discussion

Transgenes are heavily used in RNAi research, both as markers of silencing and also as silencing inducers. Due to the ease of use of transgenic markers and lack of substantial phenotypic changes in number of RNAi mutants, transgenes play a major role in studying RNAi and so far they were a key component in all forward genetic screens attempting to find genes involved in RNAi (Elmayan et al. 1998; Brodersen et al. 2008; Eamens et al. 2008; He et al. 2009; Greenberg et al. 2011; Eun et al. 2012). However, such systems are usually based on stably integrated transgenes with constitutive expression for silencing induction. This approach has its advantages, but it usually allows only to observe the final state of silencing (or some later stages). Alternative approaches exist which make use of viral infection, *Agrobacterium* infiltration or recently developed direct application of synthetic sRNAs (Ruiz et al. 1998; Dadami et al. 2014; Dalakouras et al. 2016). In these cases the silencing starts with the infection and it can be monitored throughout the whole course. But the antiviral RNAi is quite specific and the RNAi components are highly redundant in this process (Deleris et al. 2006), also the infection will not affect the plant homogeneously. These issues can be

addressed by stably integrated transgenes with inducible expression. Although this approach has its own technical difficulties, few such studies have been done, all on whole plant level using inducible inverted repeat as silencer (Chen et al. 2003; Lo et al. 2005; Wielopolska et al. 2005).

To address the aims and questions postulated in the chapter Aims of the Thesis we developed a model system composed from the following components: tobacco BY-2 cell line – to have highly homogeneous cells, which can be easily treated with chemicals; *GFP* as a target for silencing – to easily monitor the silencing even in large populations; estradiol-inducible silencers (precursors for sRNAs) – to be able to activate the silencer transcription at specific time, modulate the strength of transcription and to have the ability to switch the silencing off by washing the estradiol away (Zuo et al. 2000). We also choose to use different types of silencers to be able to compare them and describe differences between them. The types of silencers reflected basic principles how the precursor dsRNA can be formed (see Introduction), specifically: intramolecular pairing (hairpin RNA produced from inverted repeat *GFP* silencer – *IR*), intermolecular pairing (pairing of target sense *GFP* mRNA with RNA from antisense *GFP* silencer – *AS*) and second strand synthesis by RDR (aberrant RNA produced by unterminated *GFP* silencer – *UT*; Fig. 4). To be able to compare the silencers a single transgenic GFP expressing line was selected to allow direct comparison of the silencers without being affected by any variability of the silencing target, which could possibly manifest between transgenic lines with different insertion events. This line has been stably expressing GFP for several years (Nocarova & Fischer 2009). Of course there are also some disadvantages to such a system. Plant cell cultures show large epigenetic differences compared to their parental plants (Tanurdzic et al. 2008), however, the pathways themselves work efficiently enough in cell cultures (Wang & Waterhouse 2000). One study showed that PTGS might be less efficient in cell culture (Marjanac et al. 2009), while different study saw higher efficiency for TGS in cell culture (Fojtová et al. 2003). But given the large differences in silencing efficiency between tissues of a single plant, the cell culture can hardly be considered an outlier in this respect (Marjanac et al. 2009). The problem with stably integrated transgenes for silencers is the variability caused by the positional effect of integration, effects of variable number of insertions and risk of generating loci with repeat arrangement that could affect the types of produced RNAs (Wang & Waterhouse 2000; Lechtenberg et al. 2003; Fischer et al. 2008). It has been also reported that endogenes behave differently compared to transgenes when targeted by silencing. Endogenes show higher resistance to secondary sRNAs production and RdDM (Christie et al. 2011; Vermeersch et al. 2013). Therefore transgenes might be considered too artificial for such experiments, on the other hand higher sensitivity to silencing can be viewed as advantageous, because in a system too

resistant to silencing we might not be able to observe effects of weaker silencers like the antisense construct (Chen et al. 2003). It could be also speculated that transgene behavior might resemble behavior of endogenes that are supposed to be regulated by RNAi, as opposed to endogenes that are not normally regulated by RNAi, but are often used as targets for silencing in many experiments (Christie et al. 2011; Kim et al. 2021).

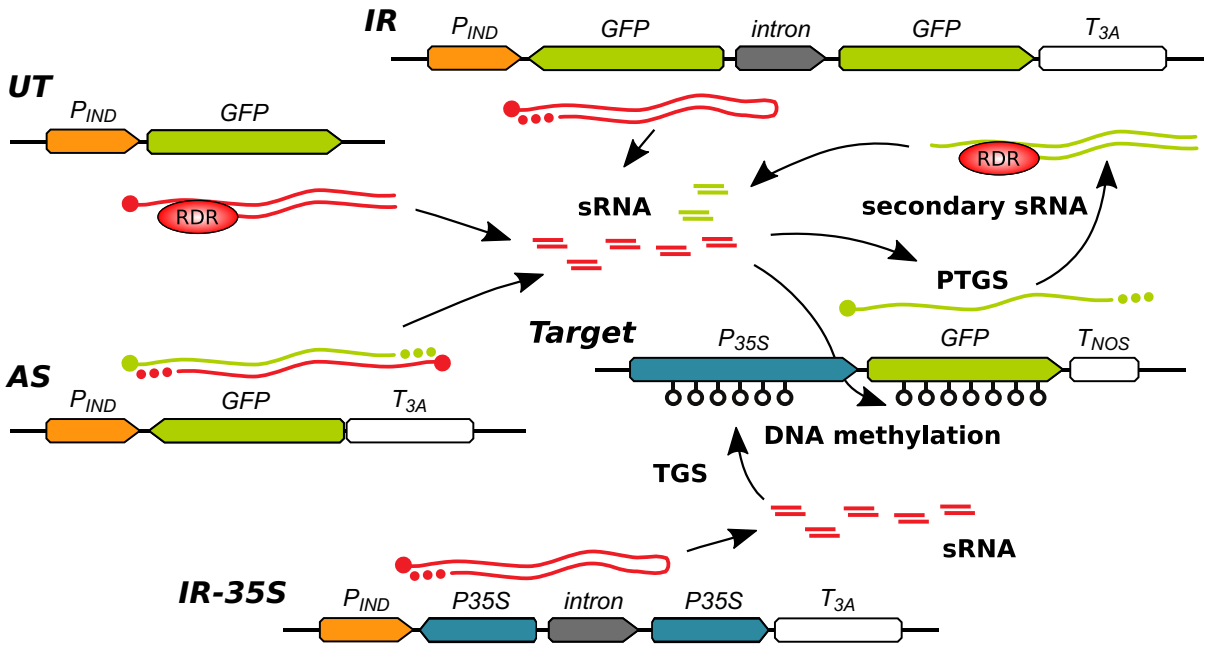


Fig. 4: Schematic representation of the mechanism of function for all the silencer variants. AS (RNA from antisense GFP silencer pairing to the target sense GFP mRNA), UT (unterminated GFP silencer producing aberrant RNA recognized by RDR), GFP IR (inverted repeat GFP silencer producing hairpin RNA), 35S IR (inverted repeat CaMV 35S silencer producing hairpin RNA for induction of TGS). The dsRNAs are processed into sRNAs which can target mRNA for degradation and possibly secondary sRNAs production (green) and DNA for methylation. P_{35S}: CaMV 35S promoter, P_{IND}: inducible promoter activated by β-estradiol; T_{NOS}: nopaline synthase terminator; T_{3A}: rbcs S 3A terminator; GFP: green fluorescent protein coding sequence; intron: intron from Solanum tuberosum PsbO gene. All T-DNAs are at the same scale.

With the described model system we could design the experiments. Our experimental workflow was as follows (Fig. 5): to supertransform GFP expressing BY-2 cell line with silencer constructs, select lines which retained florescence among the primary calli, divide them in half – one half use as control, the other expose to the inductor – this would give us the population efficiency of silencing and also allow us to select lines with efficient silencing to be analyzed further in more detail.

The parental line used for supertransformation was homogeneous, so one would expect that the primary calli derived from the transformation would be also homogeneous. However, this was not true and besides the differences among calli from single transformation, there were also differences

between the individual silencers (Attachment 1, Fig. 1). Actually the IR silencer showed massive silencing. The estradiol inducible XVE system used to regulate the expression was reported to be reliable with no leaky expression (Zuo et al. 2000; Moore et al. 2006). And indeed, although the silencer was transcribed in lines where the silencing occurred, the transcription did not originate from the inducible promoter. This silencing and its causes are described in Publication 1 (Čermák & Fischer 2018). The transcription was independent of the canonical promoters in the *T-DNA* and at least some of the transcripts originated outside of the *T-DNA*. This means that they could originate either in genomic DNA or in adjacent *T-DNAs* in case of tandem *T-DNA* insertions. We did not observe a higher occurrence of tandem *T-DNA* insertions among the spontaneously silenced lines, but we observed a generally higher number of *T-DNA* insertions among these lines. Correlation between silencing efficiency and *T-DNA* copy number of *IR* silencers was already previously observed (Wang & Waterhouse 2000). In our system the higher transcription of the spontaneously silenced lines suggests that the higher number of *T-DNA* insertions increased the chance of the read-through transcription. We believe that this transcription is initiated and/or allowed by specific chromatin status of given insertions. This is supported by the positive correlation of the sense and antisense read-through transcription and by its ability to read-through terminators – both these phenomenons could be most easily explained by the chromatin state of the *T-DNA* insertion loci. In later experiments we saw low levels of sRNAs in an uninduced (non-spontaneously silenced) IR line (Attachment 4, Fig. 5). Interestingly, after this line went through the stage of induced silencing and thereafter this silencing was turned off, these “background” levels of sRNAs disappeared. Assuming that these sRNAs were a result of read-through transcription, then it seems that this read-through transcription was stopped by chromatin changes that accompanied the silencing and/or the transient transcription of the silencer. It was previously shown that DNA methylation can have various effects on (read-through) transcription and also aberrant RNA production (Yan et al. 2016; Osabe et al. 2017; Butel et al. 2021). Transcription itself modifies chromatin and prevents cryptic transcription (Smolle & Workman 2013). In Publication 1 we analyzed and described this read-through transcription only at the *IR* silencer and sense *GFP* transgene, but there were some differences also among the remaining variants. The *UT* silencer showed significantly larger number of spontaneously silenced clones even though it was the least efficient silencer when the silencing was intentionally induced, i.e. relatively to the all other silences it performed better when spontaneously induced compared to the situation when it was intentionally induced (Attachment 1, Fig. 1; Attachment 4, Table 1). We did not analyze the spontaneous silencing in the *UT* silencer, which makes it difficult to speculate on the reason why the spontaneous silencing was more efficient in this variant.

The important outcome of Publication 1 is explanation for phenomenon known from the beginning of the RNAi research. When silencing was discovered in plants, the exact mechanism was not known. It was assumed that it could be facilitated either by RNA or by direct DNA-DNA interaction. Therefore promoterless constructs were often used as controls and in many cases silencing was observed (Van Blokland et al. 1994). Now when we know how easy it is for read-through transcripts to induce silencing, these older observations are no surprise.

Induction of posttranscriptional gene silencing

The above described spontaneous silencing probably biased the population analyses on induced silencing which followed. The IR showed lower silencing efficiency than was expected based on previous research (Wang & Waterhouse 2000; Wesley et al. 2001), probably because most of the silencing competent lines were lost due to the spontaneous silencing. Still, the IR was the most efficient silencer, while UT was the least effective, no better than control, i.e. inducible *GFP* with terminator (Attachment 4, Table 1). The low UT performance puts in question the role of aberrant

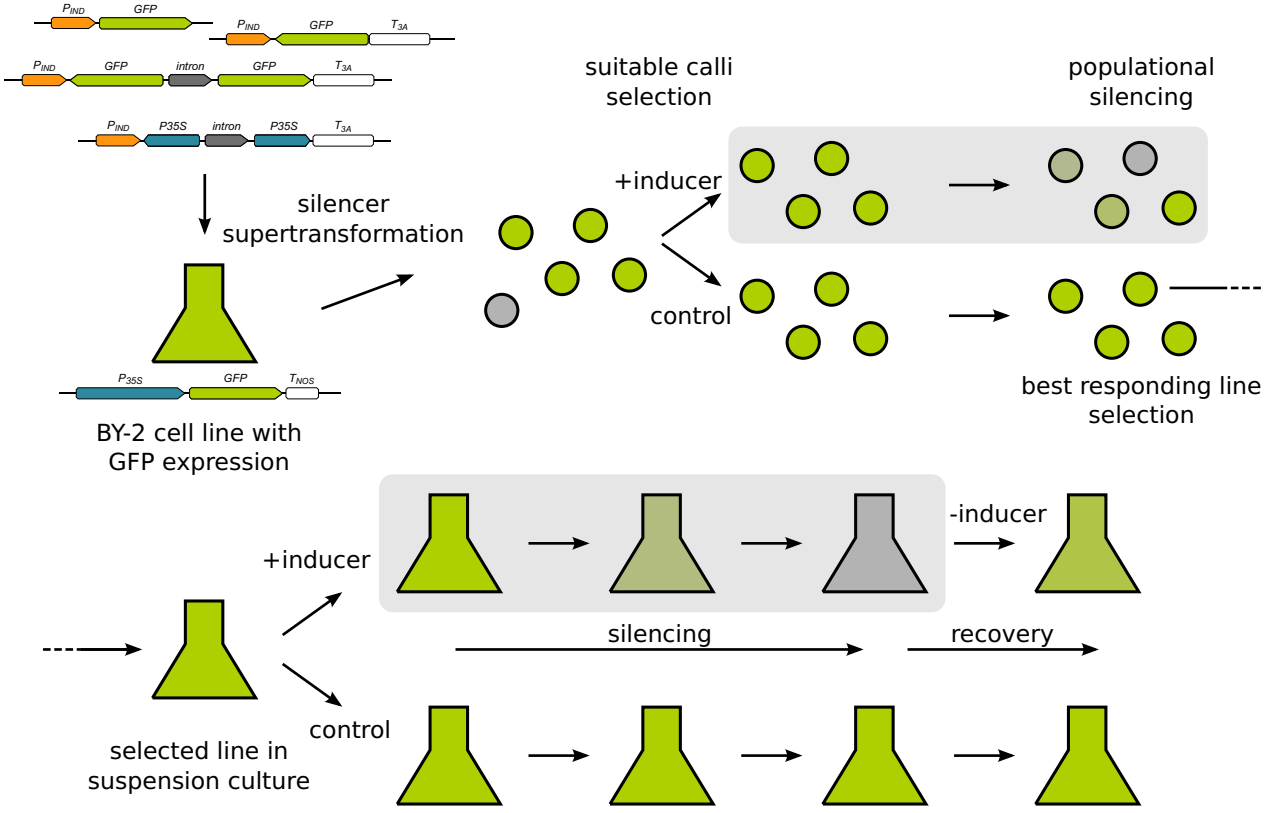


Fig. 5: Experimental workflow: BY-2 cell line with stable GFP expression was supertransformed with the silencer variants. Among the primary calli those with reasonable fluorescence were selected. Half of each calli was subjected to the inductor and the populational efficiency of silencing was evaluated. The best responding lines (calli) were selected for detailed analyses in suspension culture. Each line was exposed to the inducer for 14/10 days (for the PTGS and TGS respectively) and then the lines were monitored for another 21 days without the inductor (PTGS only). The silencing was monitored primarily as fluorescence measured by flow cytometry which was followed by transcript analyses, sRNA sequencing and DNA methylation analyses. Light gray background indicates presence of the inducer (estradiol).

RNAs in silencing, or at least the type of aberrant RNAs capable of entering the RNAi pathway (Luo & Chen 2007; Akbudak et al. 2013). It is noteworthy that the only other study, which directly compared UT silencer with other silencers, also failed to find the effect of absent terminator (Yan et al. 2006). However, the results for the UT lines were also affected by the spontaneous silencing. If we were to imagine that the *UT* silencer expression is driven by a constitutive promoter, then the UT silencing would probably be the sum of what we saw as spontaneous silencing and inducible silencing, making the *UT* silencer more efficient than control. This feature of the inducible system can be viewed as an advantage, as it discriminates against clones where the silencing is triggered by other transcripts than intended.

Out of the population of analyzed lines we selected those with strongest response – in such lines the RNAi pathways in question should work at maximum of their capacity and thus allow comparison of the silencers. The selected lines would be then subjected to thorough analyses on the course and effects of silencing for a given silencer.

As mentioned above, one of the advantages of BY-2 cell line is its homogeneity. However, such homogeneity is not absolute and besides rare instances when hybrid lines of different transgenic events are generated during transformation procedure (Nocarova & Fischer 2009), the cells can also differ in the epigenetic status of the transgene. In a previous work in our laboratory, it was shown that almost 90% of primary transgenic BY-2 lines show some sort of cell to cell variability of transgene expression, which is caused either by some heritable epigenetic state likely established short time after transformation or some instability in expression, which can have various reasons (Nocarova & Fischer 2009). In order to make sure that we work with as homogeneous cultures as possible and to understand the cell to cell variability of silencing (e.g. whether 50% decrease in fluorescence means 50% of decrease in all cells or 100% of decrease in 50% of cells), we needed to establish a method to analyze transgene expression at single cell level. After testing a few approaches, the only reliable and precise enough method turned out to be to measure fluorescence of protoplasts using flow cytometry. This method is described in Publication 2 (Klíma et al. 2019). It allowed us to analyze the behavior of individual lines in high detail (see below e.g. weak silencing induction in IR lines).

The detailed analysis of individual lines is described in Publication 4 (Čermák et al. 2020). Each cell line was exposed to the inductor for 14 days, after that it was washed out and the cell lines were monitored for an additional 21 days (Fig. 5). The fluorescence was measured at several time points. Based on the fluorescence results, samples from the potentially interesting time points were subjected to transcript analyses, sRNA sequencing and DNA methylation analyses. Besides

describing several already known or expected characteristics of the three types of silencers, we also came across several unexpected characteristics, many of which would be difficult to describe with different model systems.

The most divergent from the expected behavior were the UT lines. It should be noted that the UT lines did not form a group with uniform behavior, unlike the other silencers. It rather seemed that their behavior was affected by the number and type of *T-DNA* insertions. All the lines contained multiple *T-DNA* insertions and we could prove at least one inverted repeat and one direct tandem repeat (in three analyzed lines). It has been proposed that silencing (including transgene silencing, cossupression and antisense technology) is caused by formation of *T-DNA* inverted repeats (Stam et al. 1997; Wang & Waterhouse 2000). In our case the UT line with *T-DNAs* integrated as inverted repeat behaved like IR lines, while the remaining two UT lines behaved differently. In these we were not able to prove the presence of inverted repeats. The theory about cossupression being caused by inverted repeat arrangement was later challenged, or rather extended, by theory with aberrant RNAs that need to cross certain expression threshold (Lechtenberg et al. 2003; Luo & Chen 2007; Martínez de Alba et al. 2015). This theory was recently supported by pointing to the choice of terminator as the major cause of transgene silencing (Felippes et al. 2020). However, the aberrant RNA theory also does not entirely fit the behavior of the UT lines. As already mentioned they did not show higher efficiency of silencing than control, even though they likely produced higher amounts of aberrant RNAs than control (Attachment 4, Fig. S1C). Their silencing strength also was similar to the IR lines when the silencers were induced to lower levels – if the aberrant RNAs need to cross certain threshold to enter the RNAi pathway, it would be expected that the UT lines will stop working at low enough concentration of the inductor, which we did not observe. Since the two UT lines cannot form dsRNA by inter- or intramolecular pairing and it seems that their expression does not need to cross any threshold, the only other explanation is that they produce specific transcripts that can be directly targeted by some RDR. Nature of these transcripts remains unknown.

However, we saw a threshold for silencing in the IR lines, when the silencer was weakly induced, only some of the cells reacted (Attachment 4, Fig. 4). We believe that there could be a threshold in the amount of sRNAs in the cell needed to initiate silencing, to protect the cell from accidentally produced sRNAs. Not much is known about the kinetics of sRNA silencing in plants. In mammalian cells when miRNA target transcripts exceed a certain threshold they are able to escape silencing. But mammalian miRNAs work mostly through translational repression and this is achieved by titration of all functional miRNAs by too many target molecules. This threshold

disappears when the miRNA are able to catalytically slice and degrade the target molecules (Mukherji et al. 2011). Although this shows to a possibility of creating thresholds in activity of sRNAs, our threshold is different and the kinetics of plant sRNAs also differ. Therefore there has to be some other mechanism behind our observation. It has been shown in yeasts that there are mechanisms degrading low abundant sRNAs (Pisacane & Halic 2017). Also mathematical models of RNAi which account for mRNA targeting by multiple siRNAs predict a threshold of siRNA concentration which needs to be crossed for efficient silencing. This model best fits the data of RNAi in mammalian cells transfected by siRNAs (Cuccato et al. 2011). Similar processes likely act in plants and unlike mammals, plants are able to produce secondary sRNAs – a process which can further strengthen such threshold. Our observation could be theoretically just an artifact of the inducible promoter, but additional support for this threshold to be at the sRNA level is the presence of low levels of sRNAs in the uninduced IR and UT line, which did not cause any obvious silencing (Attachment 4, Fig. 5).

Taken together, these results on PTGS induction can be summarized in a following model: after transcription different RNAs can form dsRNAs with various efficiency. In case of aberrant RNAs this efficiency can be affected by RNA degradation pathways, but it is not necessary for the aberrant RNAs (at least of certain type) to saturate these pathways. The aberrant RNAs are available to the RNAi machinery at all concentrations. To prevent sporadically produced aberrant RNAs from triggering silencing, there is a threshold at the level of sRNAs, preventing rare sRNAs to induce silencing. This threshold applies not only to the RNAi triggered by aberrant RNAs but also to RNAi triggered by intra- or intermolecular pairing of RNAs. This model can be easily applied to previous observations, where silencing correlated with expression strength and mutations in RNA degradation pathways and which lead to suggestion of a model where RNAs have to first saturate RNA degradation pathways to be able to cross threshold which allows them to enter RNAi pathway (Schubert et al. 2004; Voinnet 2008; Martínez de Alba et al. 2015). The shift of the threshold from the precursor molecule (aberrant RNA) to the effector molecule (sRNA) makes this model more universal and since this threshold likely exists at the sRNA level, another threshold at the precursor RNA level is in most cases no longer necessary to explain given observations. Of course the threshold at the precursor level can still exist, especially in situations where given aberrant RNA can be very efficiently removed by RNA degradation pathways.

The most intriguing behavior of UT lines was their ability to maintain silencing. Surprisingly this silencing maintenance was not connected with transition of PTGS to TGS and the spread of methylation to the promoter. The maintenance seemed to be connected with perpetual production of

secondary sRNAs. What features of these sRNAs allow to maintain the production of secondary sRNAs is not known. Similar maintenance of silencing was observed on whole plant level using viral induction of silencing (Ruiz et al. 1998; Zhao et al. 2014). In our case we had the opportunity to compare the sRNAs from the UT line to sRNAs from the AS line. Both these silencers induced production of secondary sRNAs, this was most evident based on the 3' transitive sRNAs present in the 3' UTR of target GFP. In the UT line these transitive sRNAs dominated. However, these transitive sRNAs were almost identical between the AS and UT line – same abundance, same sizes, similar strand ratio and almost the same 5' end nucleotides (Attachment 4, Fig. S5), yet they were not able to maintain the silencing in the AS line. Grafting experiments on the systemic spreading of silencing in plants suggested that the silencing maintenance depends on the nature of the target locus. When the target locus was able to spontaneously switch on the silencing, only then it was able to maintain the silencing when it was triggered by external supply of sRNAs by grafting (Vaucheret et al. 2001). But our system used a single target which behaved differently when targeted with very similar sRNAs only from different source. There is one more line of evidence supporting the different nature of sRNAs between AS and UT lines. In the UT lines the sRNAs are able to methylate the target *GFP* but in the AS line they are not (Attachment 4, Fig. 6). Such difference can be hardly attributed to anything else than to the sRNAs. How to explain these situations – similar sRNAs targeting the same locus with these two different outcomes? The possible answer could be in localization. The aberrant RNAs do not have the features needed for nuclear export (Huang & Carmichael 1996), therefore the primary sRNAs have to be produced in the nucleus. On other hand the sRNAs induced by the AS silencer can be produced in cytoplasm and if their import to nucleus would be limited, it could easily explain their lower ability to induce methylation. This theory, however, has few issues that would need to be resolved. The first is the localization of sRNAs processing enzymes, all are localized to the nucleus and only few are also in cytoplasm (Hoffer et al. 2011; Pontes et al. 2013; Pumplin et al. 2016). The second issue is that the current model for secondary sRNAs production expects coupling of this process to translation and thus localization to cytoplasm. However, the secondary sRNAs can be probably produced also in the nucleus as all the necessary components are there and secondary sRNAs have been detected also from pre-mRNAs (Hoffer et al. 2011). Given the different nature of cytoplasm and nucleus in respect to the RNAi pathways, one would expect that the same sequence should produce a different spectrum of sRNAs when processed in these compartments, but we observed striking similarity in the sRNAs between AS and UT lines. More research on the compartmentalization of RNAi pathways is needed.

DNA methylation and transcriptional gene silencing

Besides analyzing the effects of RNAi on target transcripts, we also looked at target chromatin, namely DNA methylation. As sRNAs have the ability to induce DNA methylation, genes targeted for PTGS are usually also methylated. However this methylation has no known effects on PTGS or gene expression (Wang & Waterhouse 2000; Taochy et al. 2019). In our system the silencing can be also accompanied by DNA methylation. This DNA methylation mostly correlated with the strength of silencing except the already mentioned UT lines, which seemed to be more potent in methylating DNA relative to the number of sRNAs they produced. Not only was there a difference in the sRNA ability to methylate the DNA, but the DNA sequences also differed in their ability to be methylated. In each line the *GFP* sequence was present in at least two copies – as the target sequence and as a part of the silencer. In cases when the RdDM was fully induced (IR and UT lines) both these sequences were methylated more or less equally, however, in case of weaker RdDM induction (AS lines and IR with low expression of the silencer), only the silencer sequence was methylated. Different sensitivity of the identical sequences to DNA methylation has been already observed before for endogenes and transgenes (Vermeersch et al. 2013). And also among transgenes as a result of insertions in different positions (Fischer et al. 2008). Although we had only one target in one locus, we had multiple silencers in multiple loci which were always more sensitive to DNA methylation than the target. Therefore it seems that this sensitivity is probably connected with the particular *T-DNA* and not a result of the positional effect of the insertion. In our system the silencer is only transiently transcribed compared to the GFP expressed constitutively for years. Such a situation will likely result in different epigenetic status at those loci that can affect the introduction of new epigenetic modifications. Using the transcription inducible system also for the target could provide an opportunity to study such phenomena.

The DNA methylation never spread beyond the *GFP* transcription unit. In the IR lines it was distributed all along the primary target sequence and there was no correlation between the sRNAs hotspots and methylation hotspots, although the differences in methylation along the sequence were not as big as in the case of sRNAs. Therefore the RdDM was likely fully induced along the sequence, but it had different efficiency in methylating individual regions of the sequence. In the UT lines the induced methylation was at the 3’ end correlating with the amount of sRNAs. The DNA methylation is known to be able to spread from the primary target sequence (Daxinger et al. 2009; Ahmed et al. 2011). If this methylation spreads to a promoter it will result in a shift from PTGS to TGS (Fojtová et al. 2003; Marí-Ordóñez et al. 2013). Such a situation can be desirable for loci that should be silenced long term, like transposons (Marí-Ordóñez et al. 2013; Fultz & Slotkin

2017). We did not observe such spreading probably because of too short time of silencing induction. It would be intriguing also to describe the dynamics of this process. But the time necessary to observe spreading of methylation would probably require months or few generations of plants (Fojtová et al. 2003; Nocarova et al. 2010; Catoni et al. 2013; Marí-Ordóñez et al. 2013; Weinhold et al. 2013).

To better understand the dynamics of TGS, we chose to induce TGS directly. These results are described in Publication 3 (Přibylová et al. 2019). The system was the same as for PTGS induction, we used inverted repeat made with 5' fragment of *CaMV 35S* promoter. The population efficiency of silencing was only slightly lower than in the *IR* induced PTGS, however, in this case not affected by spontaneous silencing (Motylová 2015). After induction in the selected lines the reaction was rapid, the first significant silencing was apparent in 12 hours and full silencing was reached in two days in the best responding lines. Comparing the IR induced PTGS and TGS, the PTGS was 50% faster in affecting the phenotype (fluorescence) and three times faster at decreasing the mRNA transcript. Unlike TGS, PTGS increases the transcript turnover, which itself is probably quite fast given how closely the decreasing transcript levels were able to follow the increasing methylation during TGS (Attachment 3, Fig. 2). Same as the decreasing transcript levels, the first methylation appeared in 12 hours and reached maximum in two days. There were differences in the dynamics of the individual cytosine contexts. The CHH methylation peaked at two days of silencing while the other two contexts kept increasing further, likely due to involvement of other methyltransferases in their maintenance. These dynamics are affected by cell division as the DNA methylation has to be reestablished at the newly synthesized DNA strands. Considering that BY-2 doubles at about every 20 hours, the RdDM is able to methylate already silenced locus at much faster rate than the naive locus. But the mere presence of methylated CG had no effect on *de novo* methylation as evident from one of the lines with CG methylation in the promoter before the start of the experiment. Taken together the sRNAs in TGS act very fast, with some delay at the start of methylation and somewhat more gradual accumulation of methylcytosines, possibly hampered by the ongoing cell division. The dynamics of methylation of transcribed loci during PTGS differ a little bit, the methylation of untranscribed promoter can be much faster – more than twice compared to the *GFP IR* silencer and more than three times compared to the target *GFP* (Attachment 4, Fig. S6). This again points to a different sensitivity of loci to methylation probably as a result of transcription as discussed before.

It should be also pointed out that the two *IRs* differed in the spectrum of produced sRNAs. The P35S IR lines produced mainly 21 nt siRNA with very low levels of 24 nt sRNA. The GFP IR line produced mainly 22 nt sRNAs with relatively higher levels of 24 nt sRNAs. The difference in the

amount of produced 24 nt sRNAs further highlights the difference in the speed of methylation between the two *IRs* and the impact of the nature of the target locus at its methylation. The reason for the different composition of produced sRNAs is not known. The sequences of the two *IRs* differed, the *GFP IR* also had the opportunity to produce secondary sRNAs at the target, although the lack of transitive sRNAs at the *IR* silenced *GFP* suggests that there were no or very low levels of secondary sRNAs from the target. Also we sequenced only one GFP IR line, which has undergone recombination during transformation that resulted in the absence of intron between the *IR* arms – therefore unlike the other *IRs* this one could not be spliced during transcription.

The RdDM was not the only mechanism that maintained the established DNA methylation. MET1 was involved at maintaining CG methylation as evident from the persistence of methylated CG in the absence of sRNAs (Attachment 4, Fig. 6) (Aufsatz et al. 2004). While the CMT3 maintained the CHG methylation as evident from the higher methylation of CWG sequences (W = A or T) than in CCG sequences – typical hallmark of CMT3 activity (Gouil & Baulcombe 2016). Interestingly the methylation of CWG sequences reached 100%. This contradicts the conclusions of Borges et al. 2020 who suggested that CMT3 activity lags after cell division resulting in hemimethylated CWG sites in dividing cells. Our results suggest that such behavior is not a general CMT3 feature, but it is rather specific to given conditions or species. Unlike CG methylation the CHG methylation was not maintained in the *GFP* coding sequence. This is not surprising because the CMT3 activity depends on H3K9me2 modification which is actively removed from transcribed regions (Saze et al. 2008). The fate of DNA methylation in individual contexts in the promoter region in absence of sRNAs was not analyzed and is a subject of ongoing experiments (Přibylová, unpublished results).

Conclusions

We developed a system to study the dynamics of RNAi induced by various silencers (triggers of sRNA production). The system was based on BY-2 cell line with stable expression of *GFP* reporter under the control of *CaMV 35S* promoter. The silencers were under the control of an estradiol inducible promoter. We used four different silencers: *AS* (RNA from antisense *GFP* silencer pairing to the target sense *GFP* mRNA), *UT* (unterminated *GFP* silencer producing aberrant RNA recognized by RDR), *GFP IR* (inverted repeat *GFP* silencer producing hairpin RNA), *35S IR* (inverted repeat *CaMV 35S* silencer producing hairpin RNA for induction of TGS; Fig. 4). With this system we made the following observations and conclusions:

The silencer differed in several quantitative characteristics – efficiency: *GFP IR* > *35S IR* > *AS* > *UT*; strength: *35S IR* ≥ *GFP IR* > *AS* ≥ *UT*; speed: *GFP IR* > *35S IR* > *AS* ≥ *UT*; sensitivity: *GFP IR* > *UT* > *AS* (n.d. for *35S IR*); amount of produced sRNAs: *35S IR* ≥ *GFP IR* > *AS* > *UT*; proportion of target DNA methylation: *GFP IR* ≥ *35S IR* > *UT* > *AS*.

Compared to the other silencers, the *AS* showed sRNA-strand bias in favor of the reverse strand sRNAs. It also showed asymmetry in its ability to methylate DNA, with the silencer sequence being more sensitive to methylation while the target sequence was resistant. Similar phenomenon was also observed with low induced *GFP IR*.

UT along with *AS* produced high amounts of secondary and transitive sRNAs. In the UT line these sRNAs persisted without the induction and they were able to maintain the silencing. The behavior of the UT lines also puts in question the mechanisms of RNAi induced by aberrant RNAs. The impact of the missing terminator was questionable in our system. We also did not observe any expression threshold for the aberrant RNAs, which was previously suggested as a result of the need to saturate RNA degradation pathways.

As expected the *GFP IR* silencer topped all the quantitative characteristics. But in our system it also allowed us to compare strongly and weakly induced RNAi. At very weak induction we observed binary silencing at individual cell level (cells were either silenced or not), suggesting a certain sRNAs threshold had to be exceeded. If confirmed, such threshold could have wide implications for interpretation of data on plant RNAi.

The DNA methylation induced by *35S IR* was very fast (first methylation in 12 hours, full practically in two days), resulting in quick TGS. The dynamics of DNA methylation differed in individual cytosine contexts reflecting involvement of different maintenance DNA methyltransferases after establishment of primary methylation.

Besides inducible silencing we also observed spontaneous silencing – silencing caused by read-through transcription from outside the *T-DNA*. The most affected was the *GFP IR* silencer, likely because it was the most efficient in silencing.

References

Agius, F., Kapoor, A. & Zhu, J.-K., 2006. Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proceedings of the National Academy of Sciences*, 103(31), 11796–11801.

Ahmed, I., Sarazin, A., Bowler, C., Colot, V. & Quesneville, H., 2011. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Research*, 39(16), 6919–6931.

Akbudak, M.A., Nicholson, S.J. & Srivastava, V., 2013. Suppression of Arabidopsis genes by terminator-less transgene constructs. *Plant Biotechnology Reports*, 7(4), 415–424.

Arribas-Hernández, L., Marchais, A., Poulsen, C., Haase, B., Hauptmann, J., Benes, V., Meister, G. & Brodersen, P., 2016. The Slicer Activity of ARGONAUTE1 Is Required Specifically for the Phasing, Not Production, of Trans-Acting Short Interfering RNAs in Arabidopsis. *The Plant Cell*, 28(7), 1563–1580.

Aufsatz, W., Mette, M., Matzke, A. & Matzke, M., 2004. The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Molecular Biology*, 54(6), 793–804.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M. & Matzke, A.J.M., 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *The EMBO journal*, 21(24), 6832–6841.

Axtell, M.J. & Meyers, B.C., 2018. Revisiting criteria for plant miRNA annotation in the era of big data. *The Plant Cell*, tpc.00851.2017.

Baeg, K., Iwakawa, H. & Tomari, Y., 2017. The poly(A) tail blocks RDR6 from converting self mRNAs into substrates for gene silencing. *Nature Plants*, 3(4), 17036.

Baumberger, N. & Baulcombe, D.C., 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences*, 102(33), 11928–11933.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M. & Jacobsen, S.E., 2008. Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana. *PLoS ONE*, 3(9), e3156.

Bhattacharjee, S., Roche, B. & Martienssen, R.A., 2019. RNA-induced initiation of transcriptional silencing (RITS) complex structure and function. *RNA Biology*, 16(9), 1133–1146.

Böhmdorfer, G., Sethuraman, S., Rowley, M.J., Krzyszton, M., Rothi, M.H., Bouzit, L. & Wierzbicki, A.T., 2016. Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *eLife*, 5, e19092.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M. & Benning, C., 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *The EMBO Journal*, 17(1), 170–180.

Borg, M., Jiang, D. & Berger, F., 2021. Histone variants take center stage in shaping the epigenome. *Current Opinion in Plant Biology*, 61, 101991.

Borges, F., Donoghue, M.T.A., LeBlanc, C., Wear, E.E., Tanurdžić, M., Berube, B., Brooks, A., Thompson, W.F., Hanley-Bowdoin, L. & Martienssen, R.A., 2020. Loss of Small-RNA-Directed DNA Methylation in the Plant Cell Cycle Promotes Germline Reprogramming and Somaclonal Variation. *Current Biology*.

Borges, F. & Martienssen, R.A., 2015. The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, advance online publication.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L. & Voinnet, O., 2008. Widespread Translational Inhibition by Plant miRNAs and siRNAs. *Science*, 320(5880), 1185–1190.

Butel, N., Yu, A., Le Masson, I., Borges, F., Elmayan, T., Taochy, C., Gursanscky, N.R., Cao, J., Bi, S., Sawyer, A., et al., 2021. Contrasting epigenetic control of transgenes and endogenous genes promotes post-transcriptional transgene silencing in Arabidopsis. *Nature Communications*, 12(1), 2787.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M. & Jacobsen, S.E., 2003. Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. *Current Biology*, 13(24), 2212–2217.

Cao, X. & Jacobsen, S.E., 2002. Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing. *Current Biology*, 12(13), 1138–1144.

Catoni, M., Lucioli, A., Doblas-Ibáñez, P., Accotto, G.P. & Vaira, A.M., 2013. From immunity to susceptibility: virus resistance induced in tomato by a silenced transgene is lost as TGS overcomes PTGS. *The Plant Journal*, 75(6), 941–953.

Cecere, G. & Grishok, A., 2014. A nuclear perspective on RNAi pathways in metazoans. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(3), 223–233.

Čermák, V. & Fischer, L., 2018. Pervasive read-through transcription of T-DNAs is frequent in tobacco BY-2 cells and can effectively induce silencing. *BMC Plant Biology*, 18(1), 252.

Čermák, V., Tyč, D., Přibylová, A. & Fischer, L., 2020. Unexpected variations in posttranscriptional gene silencing induced by differentially produced dsRNAs in tobacco cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(11), 194647.

Chen, S., Hofius, D., Sonnewald, U. & Börnke, F., 2003. Temporal and spatial control of gene silencing in transgenic plants by inducible expression of double-stranded RNA. *The Plant Journal*, 36(5), 731–740.

Christie, M., Croft, L.J. & Carroll, B.J., 2011. Intron splicing suppresses RNA silencing in Arabidopsis. *The Plant Journal*, 68(1), 159–167.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. & Jacobsen, S.E., 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452(7184), 215–219.

Cuccato, G., Polynikis, A., Siciliano, V., Graziano, M., di Bernardo, M. & di Bernardo, D., 2011. Modeling RNA interference in mammalian cells. *BMC Systems Biology*, 5(1), 19.

Dadami, E., Dalakouras, A., Zwiebel, M., Krczal, G. & Wassenegger, M., 2014. An endogene-resembling transgene is resistant to DNA methylation and systemic silencing. *RNA Biology*, 11(7), 934–941.

Dalakouras, A., Wassenegger, M., McMillan, J.N., Cardoza, V., Maegele, I., Dadami, E., Runne, M., Krczal, G. & Wassenegger, M., 2016. Induction of Silencing in Plants by High-Pressure Spraying of In vitro-Synthesized Small RNAs. *Frontiers in Plant Science*, 7.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J.M. & Matzke, M., 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *The EMBO Journal*, 28(1), 48–57.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C. & Voinnet, O., 2006. Hierarchical Action and Inhibition of Plant Dicer-Like Proteins in Antiviral Defense. *Science*, 313(5783), 68–71.

Diaz-Pendon, J.A., Li, F., Li, W.-X. & Ding, S.-W., 2007. Suppression of Antiviral Silencing by Cucumber Mosaic Virus 2b Protein in Arabidopsis Is Associated with Drastically Reduced Accumulation of Three Classes of Viral Small Interfering RNAs. *Plant Cell*, 19(6), 2053–2063.

Duan, C.-G., Zhang, H., Tang, K., Zhu, X., Qian, W., Hou, Y.-J., Wang, B., Lang, Z., Zhao, Y., Wang, X., et al., 2015. Specific but interdependent functions for Arabidopsis AGO4 and AGO6 in RNA-directed DNA methylation. *The EMBO Journal*, 34(5), 581–592.

Eamens, A., Vaistij, F.E. & Jones, L., 2008. NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. *The Plant Journal*, 55(4), 596–606.

Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.-B. & Waterhouse, P.M., 2009. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA*, 15(12), 2219–2235.

Elmayan, T., Balzergue, S., Béon, F., Bourdon, V., Daubremet, J., Guénet, Y., Mourrain, P., Palauqui, J.-C., Vernhettes, S., Vialle, T., et al., 1998. Arabidopsis Mutants Impaired in Cosuppression. *The Plant Cell*, 10(10), 1747–1757.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.-A., Jacobsen, S.E., Cooke, R. & Lagrange, T., 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes & Development*, 21(20), 2539–2544.

Enke, R.A., Dong, Z. & Bender, J., 2011. Small RNAs Prevent Transcription-Coupled Loss of Histone H3 Lysine 9 Methylation in Arabidopsis thaliana. *PLoS Genet*, 7(10), e1002350.

Eun, C., Lorkovic, Z.J., Sasaki, T., Naumann, U., Matzke, A.J.M. & Matzke, M., 2012. Use of Forward Genetic Screens to Identify Genes Required for RNA-Directed DNA Methylation in Arabidopsis thaliana. *Cold Spring Harbor Symposia on Quantitative Biology*, 77, 195–204.

Fei, Q., Xia, R. & Meyers, B.C., 2013. Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *The Plant Cell Online*, 25(7), 2400–2415.

Felippes, F.F. de, McHale, M., Doran, R.L., Roden, S., Eamens, A.L., Finnegan, E.J. & Waterhouse, P.M., 2020. The key role of terminators on the expression and post-transcriptional gene silencing of transgenes. *The Plant Journal*, 104(1), 96–112.

Fischer, U., Kuhlmann, M., Pecinka, A., Schmidt, R. & Mette, M.F., 2008. Local DNA features affect RNA-directed transcriptional gene silencing and DNA methylation. *The Plant Journal*, 53(1), 1–10.

Fojtová, M., Houdt, H.V., Depicker, A. & Kovařík, A., 2003. Epigenetic Switch from Posttranscriptional to Transcriptional Silencing Is Correlated with Promoter Hypermethylation. *Plant Physiology*, 133(3), 1240–1250.

Fultz, D. & Slotkin, R.K., 2017. Exogenous Transposable Elements Circumvent Identity-Based Silencing Permitting the Dissection of Expression Dependent Silencing. *The Plant Cell*, tpc.00718.2016.

Gallego-Bartolomé, J., Liu, W., Kuo, P.H., Feng, S., Ghoshal, B., Gardiner, J., Zhao, J.M.-C., Park, S.Y., Chory, J. & Jacobsen, S.E., 2019. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. *Cell*.

Gehring, M., Bubb, K.L. & Henikoff, S., 2009. Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting. *Science*, 324(5933), 1447–1451.

Gouil, Q. & Baulcombe, D.C., 2016. DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLOS Genetics*, 12(12), e1006526.

Greenberg, M.V.C., Ausin, I., Chan, S.W.L., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C., et al., 2011. Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics*, 6(3), 344–354.

Greenberg, M.V.C., Deleris, A., Hale, C.J., Liu, A., Feng, S. & Jacobsen, S.E., 2013. Interplay between Active Chromatin Marks and RNA-Directed DNA Methylation in Arabidopsis thaliana. *PLoS Genet*, 9(11), e1003946.

Harris, C.J., Scheibe, M., Wongpalee, S.P., Liu, W., Cornett, E.M., Vaughan, R.M., Li, X., Chen, W., Xue, Y., Zhong, Z., et al., 2018. A DNA methylation reader complex that enhances gene transcription. *Science*, 362(6419), 1182–1186.

Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H. & Baulcombe, D.C., 2010. The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *The Plant Cell Online*, 22(2), 321–334.

He, X.-J., Hsu, Y.-F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.-S. & Zhu, J.-K., 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes & Development*, 23(3), 318–330.

Herr, A.J., Molnàr, A., Jones, A. & Baulcombe, D.C., 2006. Defective RNA processing enhances

RNA silencing and influences flowering of Arabidopsis. *Proceedings of the National Academy of Sciences*, 103(41), 14994–15001.

Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., Koiwa, H., Seki, M., Shinozaki, K. & Fukuhara, T., 2005. Specific interactions between Dicer-like proteins and HYL1/DRB- family dsRNA-binding proteins in Arabidopsis thaliana. *Plant Molecular Biology*, 57(2), 173–188.

Hoffer, P., Ivashuta, S., Pontes, O., Vitins, A., Pikaard, C., Mroczka, A., Wagner, N. & Voelker, T., 2011. Posttranscriptional gene silencing in nuclei. *Proceedings of the National Academy of Sciences*, 108(1), 409–414.

Huang, Y. & Carmichael, G.C., 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Molecular and Cellular Biology*, 16(4), 1534–1542.

Ingouff, M., Selles, B., Michaud, C., Vu, T.M., Berger, F., Schorn, A.J., Autran, D., Durme, M.V., Nowack, M.K., Martienssen, R.A., et al., 2017. Live-cell analysis of DNA methylation during sexual reproduction in Arabidopsis reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes & Development*, 31(1), 72–83.

Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., et al., 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature*, 507(7490), 124–128.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L. & Richards, E.J., 2003. Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics*, 163(3), 1109–1122.

Karlowski, W.M., Zielezinski, A., Carrère, J., Pontier, D., Lagrange, T. & Cooke, R., 2010. Genome-wide computational identification of WG/GW Argonaute-binding proteins in Arabidopsis. *Nucleic Acids Research*, 38(13), 4231–4245.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A. & Carrington, J.C., 2007. Genome-Wide Profiling and Analysis of Arabidopsis siRNAs. *PLoS Biology*, 5(3), e57.

Kim, E.Y., Wang, L., Lei, Z., Li, H., Fan, W. & Cho, J., 2021. Ribosome stalling and SGS3 phase separation prime the epigenetic silencing of transposons. *Nature Plants*, 7(3), 303–309.

Klíma, P., Čermák, V., Srba, M., Müller, K., Petrášek, J., Šonka, J., Fischer, L. & Opatrný, Z., 2019. Plant Cell Lines in Cell Morphogenesis Research: From Phenotyping to -Omics. In F. Cvrčková & V. Žárský, eds. *Plant Cell Morphogenesis: Methods and Protocols*. Methods in Molecular Biology. New York, NY: Springer New York, pp. 367–376.

Krzyszton, M., Zakrzewska-Placzek, M., Kwasnik, A., Dojer, N., Karlowski, W. & Kufel, J., 2018. Defective XRN3-mediated transcription termination in Arabidopsis affects the expression of protein-coding genes. *The Plant Journal*, 93(6), 1017–1031.

Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J. & Jacobsen, S.E., 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*, 498(7454), 385–389.

Law, J.A. & Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), 204–220.

Lechtenberg, B., Schubert, D., Forsbach, A., Gils, M. & Schmidt, R., 2003. Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *The Plant Journal*, 34(4), 507–517.

Lee, C.H. & Carroll, B.J., 2018. Evolution and Diversification of Small RNA Pathways in Flowering Plants. *Plant and Cell Physiology*, 59(11), 2169–2187.

Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X., 2005. Methylation Protects miRNAs and siRNAs from a 3'-End Uridylation Activity in Arabidopsis. *Current Biology*, 15(16), 1501–1507.

Li, S., Le, B., Ma, X., Li, S., You, C., Yu, Y., Zhang, B., Liu, L., Gao, L., Shi, T., et al., 2016. Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. *eLife*, 5.

Li, X., Harris, C.J., Zhong, Z., Chen, W., Liu, R., Jia, B., Wang, Z., Li, S., Jacobsen, S.E. & Du, J., 2018. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences*, 201809841.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. & Ecker, J.R., 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell*, 133(3), 523–536.

Liu, C., Lu, F., Cui, X. & Cao, X., 2010. Histone Methylation in Higher Plants. *Annual Review of Plant Biology*, 61(1), 395–420.

Liu, W., Duttke, S.H., Hetzel, J., Groth, M., Feng, S., Gallego-Bartolome, J., Zhong, Z., Kuo, H.Y., Wang, Z., Zhai, J., et al., 2018. RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis. *Nature Plants*, 1.

Liu, X., Yu, C.-W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y. & Wu, K., 2012. HDA6 Directly Interacts with DNA Methyltransferase MET1 and Maintains Transposable Element Silencing in Arabidopsis. *Plant Physiology*, 158(1), 119–129.

Lo, C., Wang, N. & Lam, E., 2005. Inducible double-stranded RNA expression activates reversible transcript turnover and stable translational suppression of a target gene in transgenic tobacco. *FEBS Letters*, 579(6), 1498–1502.

Luo, J. & Hall, B., 2007. A Multistep Process Gave Rise to RNA Polymerase IV of Land Plants. *Journal of Molecular Evolution*, 64(1), 101–112.

Luo, Z. & Chen, Z., 2007. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. *The Plant Cell*, 19(3), 943–958.

MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D. & Doudna, J.A., 2006. Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*, 311(5758), 195–198.

Manavella, P.A., Koenig, D. & Weigel, D., 2012. Plant Secondary siRNA Production Determined by microRNA-Duplex Structure. *Proceedings of the National Academy of Sciences*, 109(7),

2461–2466.

Marasco, M., Li, W., Lynch, M. & Pikaard, C.S., 2017. Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination. *Nucleic Acids Research*.

Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J. & Waterhouse, P.M., 2006. The evolution and diversification of Dicers in plants. *FEBS Letters*, 580(10), 2442–2450.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. & Voinnet, O., 2013. Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics*, 45(9), 1029–1039.

Marjanac, G., Karimi, M., Naudts, M., Beeckman, T., Depicker, A. & De Buck, S., 2009. Gene silencing induced by hairpin or inverted repeated sense transgenes varies among promoters and cell types. *New Phytologist*, 184(4), 851–864.

Martínez de Alba, A.E., Moreno, A.B., Gabriel, M., Mallory, A.C., Christ, A., Bounon, R., Balzergue, S., Aubourg, S., Gautheret, D., Crespi, M.D., et al., 2015. In plants, decapping prevents RDR6-dependent production of small interfering RNAs from endogenous mRNAs. *Nucleic Acids Research*, gkv119.

Matzke, M.A. & Matzke, A.J.M., 2004. Planting the Seeds of a New Paradigm. *PLoS Biology*, 2(5), e133.

McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N. & Slotkin, R.K., 2015. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *The EMBO Journal*, 34(1), 20–35.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., et al., 2008. Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5’ Terminal Nucleotide. *Cell*, 133(1), 116–127.

Mishra, V., Singh, J., Wang, F., Zhang, Y., Fukudome, A., Trinidad, J.C., Takagi, Y. & Pikaard, C.S., 2021. Assembly of a dsRNA synthesizing complex: RNA-DEPENDENT RNA POLYMERASE 2 contacts the largest subunit of NUCLEAR RNA POLYMERASE IV. *Proceedings of the National Academy of Sciences*, 118(13).

Mlotshwa, S., Pruss, G.J., Peragine, A., Endres, M.W., Li, J., Chen, X., Poethig, R.S., Bowman, L.H. & Vance, V., 2008. DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis. *PLoS ONE*, 3(3), e1755.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E. & Carrington, J.C., 2008. Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell*, 133(1), 128–141.

Moore, I., Samalova, M. & Kurup, S., 2006. Transactivated and chemically inducible gene expression in plants. *The Plant Journal*, 45(4), 651–683.

Moreno, A.B., Alba, A.E.M. de, Bardou, F., Crespi, M.D., Vaucheret, H., Maizel, A. & Mallory,

A.C., 2013. Cytoplasmic and nuclear quality control and turnover of single-stranded RNA modulate post-transcriptional gene silencing in plants. *Nucleic Acids Research*, 41(8), 4699–4708.

Mosher, R.A., Schwach, F., Studholme, D. & Baulcombe, D.C., 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proceedings of the National Academy of Sciences*, 105(8), 3145–3150.

Motylová, Š., 2015. Dynamika a mechanismus umlčování reportérového genu pro GFP v závislosti na aktivitě RDR6 a způsobu indukce RNA interference v buněčné linii tabáku BY-2.

Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A. & van Oudenaarden, A., 2011. MicroRNAs can generate thresholds in target gene expression. *Nature Genetics*, 43(9), 854–859.

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H. & Fukuhara, T., 2014. Distinct substrate specificities of Arabidopsis DCL3 and DCL4. *Nucleic Acids Research*, 42(3), 1845–1856.

Nagata, T., Nemoto, Y. & Hasezawa, S., 1992. Tobacco BY-2 cell line as the “HeLa” cell in the cell biology of higher plants. *International Review of Cytology*, 132, 1–30.

Nocarova, E. & Fischer, L., 2009. Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. *BMC Plant Biology*, 9, 44.

Nocarova, E., Opatrny, Z. & Fischer, L., 2010. Successive silencing of tandem reporter genes in potato (Solanum tuberosum) over 5 years of vegetative propagation. *Annals of Botany*, 106(4), 565–572.

O’Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A. & Ecker, J.R., 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, 165(5), 1280–1292.

Osabe, K., Harukawa, Y., Miura, S. & Saze, H., 2017. Epigenetic Regulation of Intronic Transgenes in Arabidopsis. *Scientific Reports*, 7.

Panda, K., McCue, A.D. & Slotkin, R.K., 2020. Arabidopsis RNA Polymerase IV generates 21–22 nucleotide small RNAs that can participate in RNA-directed DNA methylation and may regulate genes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1795), 20190417.

Parent, J.-S., Bouteiller, N., Elmayan, T. & Vaucheret, H., 2015a. Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *The Plant Journal*, 81(2), 223–232.

Parent, J.-S., Cahn, J., Herridge, R.P., Grimanelli, D. & Martienssen, R.A., 2021. Small RNAs guide histone methylation in Arabidopsis embryos. *Genes & Development*.

Parent, J.-S., Jauvion, V., Bouché, N., Béclin, C., Hachet, M., Zytnicki, M. & Vaucheret, H., 2015b. Post-transcriptional gene silencing triggered by sense transgenes involves uncapped antisense RNA and differs from silencing intentionally triggered by antisense transgenes. *Nucleic Acids Research*, 43(17), 8464–8475.

Paszkowski, J. & Grossniklaus, U., 2011. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Current Opinion in Plant Biology*, 14(2), 195–203.

Pisacane, P. & Halic, M., 2017. Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi. *Nature Communications*, 8, 15332.

Polydore, S. & Axtell, M.J., 2018. Analysis of RDR1/RDR2/RDR6-independent small RNAs in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of short interfering RNA loci. *The Plant Journal*, 94(6), 1051–1063.

Pontes, O., Vitins, A., Ream, T.S., Hong, E., Pikaard, C.S. & Costa-Nunes, P., 2013. Intersection of Small RNA Pathways in Arabidopsis thaliana Sub-Nuclear Domains. *PLoS ONE*, 8(6), e65652.

Pontier, D., Picart, C., Roudier, F., Garcia, D., Lahmy, S., Azevedo, J., Alart, E., Laudié, M., Karlowski, W.M., Cooke, R., et al., 2012. NERD, a Plant-Specific GW Protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in Arabidopsis. *Molecular Cell*, 48(1), 121–132.

Přibylová, A., Čermák, V., Tyč, D. & Fischer, L., 2019. Detailed insight into the dynamics of the initial phases of de novo RNA-directed DNA methylation in plant cells. *Epigenetics & Chromatin*, 12(1), 54.

Pumplin, N., Sarazin, A., Jullien, P.E., Bologna, N.G., Oberlin, S. & Voinnet, O., 2016. DNA Methylation Influences the Expression of DICER-LIKE4 Isoforms, Which Encode Proteins of Alternative Localization and Function. *The Plant Cell*, 28(11), 2786–2804.

Reis, R.S., Hart-Smith, G., Eamens, A.L., Wilkins, M.R. & Waterhouse, P.M., 2015. MicroRNA Regulatory Mechanisms Play Different Roles in Arabidopsis. *Journal of Proteome Research*, 14(11), 4743–4751.

Ronemus, M., Vaughn, M.W. & Martienssen, R.A., 2006. MicroRNA-Targeted and Small Interfering RNA-Mediated mRNA Degradation Is Regulated by Argonaute, Dicer, and RNA-Dependent RNA Polymerase in Arabidopsis. *Plant Cell*, 18(7), 1559–1574.

Ruiz, M.T., Voinnet, O. & Baulcombe, D.C., 1998. Initiation and Maintenance of Virus-Induced Gene Silencing. *The Plant Cell*, 10(6), 937–946.

Saze, H., Shiraishi, A., Miura, A. & Kakutani, T., 2008. Control of Genic DNA Methylation by a jmjC Domain-Containing Protein in Arabidopsis thaliana. *Science*, 319(5862), 462–465.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S. & Schmidt, R., 2004. Silencing in Arabidopsis T-DNA Transformants: The Predominant Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. *The Plant Cell*, 16(10), 2561–2572.

Sidorenko, L.V., Lee, T., Woosley, A., Moskal, W.A., Bevan, S.A., Merlo, P.A.O., Walsh, T.A., Wang, X., Weaver, S., Glancy, T.P., et al., 2017. GC-rich coding sequences reduce transposon-like, small RNA-mediated transgene silencing. *Nature Plants*, 1.

Smolle, M. & Workman, J.L., 2013. Transcription-associated histone modifications and cryptic transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(1), 84–97.

Stam, M., De Bruin, R., Kenter, S., Van Der Hoorn, R.A.L., Van Blokland, R., Mol, J.N.M. & Kooter, J.M., 1997. Post-transcriptional silencing of chalcone synthase in Petunia by inverted transgene repeats. *The Plant Journal*, 12(1), 63–82.

Stoddard, C.I., Feng, S., Campbell, M.G., Liu, W., Wang, H., Zhong, X., Bernatavichute, Y., Cheng, Y., Jacobsen, S.E. & Narlikar, G.J., 2018. A Nucleosome Bridging Mechanism for Activation of a Maintenance DNA Methyltransferase. *Molecular Cell*.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V. & Jacobsen, S.E., 2013. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. *Cell*, 152(1–2), 352–364.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. & Watanabe, Y., 2008. The Mechanism Selecting the Guide Strand from Small RNA Duplexes is Different Among Argonaute Proteins. *Plant Cell Physiology*, 49(4), 493–500.

Tang, K., Lang, Z., Zhang, H. & Zhu, J.K., 2016. The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nature plants*, 2(11), 16169.

Tanurdzic, M., Vaughn, M.W., Jiang, H., Lee, T.-J., Slotkin, R.K., Sosinski, B., Thompson, W.F., Doerge, R.W. & Martienssen, R.A., 2008. Epigenomic Consequences of Immortalized Plant Cell Suspension Culture. *PLoS Biol*, 6(12), e302.

Taochy, C., Yu, A., Bouché, N., Bouteiller, N., Elmayan, T., Dressel, U., Carroll, B.J. & Vaucheret, H., 2019. Post-transcriptional gene silencing triggers dispensable DNA methylation in gene body in Arabidopsis. *Nucleic Acids Research*.

Tsuzuki, M., Sethuraman, S., Coke, A.N., Rothi, M.H., Boyle, A.P. & Wierzbicki, A.T., 2020. Broad noncoding transcription suggests genome surveillance by RNA polymerase V. *Proceedings of the National Academy of Sciences*.

Van Blokland, R., Van der Geest, N., Mol, J.N.M. & Kooter, J.M., 1994. Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. *The Plant Journal*, 6(6), 861–877.

Vaucheret, H., Béclin, C. & Fagard, M., 2001. Post-transcriptional gene silencing in plants. *Journal of Cell Science*, 114(17), 3083–3091.

Vermeersch, L., De Winne, N., Nolf, J., Bleys, A., Kovařík, A. & Depicker, A., 2013. Transitive RNA silencing signals induce cytosine methylation of a transgenic but not an endogenous target. *The Plant Journal*, 74(5), 867–879.

Voinnet, O., 2008. Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends in Plant Science*, 13(7), 317–328.

Wang, M.-B. & Waterhouse, P.M., 2000. High-efficiency silencing of a β-glucuronidase gene in rice is correlated with repetitive transgene structure but is independent of DNA methylation. *Plant Molecular Biology*, 43(1), 67–82.

Weinhold, A., Kallenbach, M. & Baldwin, I.T., 2013. Progressive 35S promoter methylation increases rapidly during vegetative development in transgenic Nicotiana attenuata plants. *BMC Plant Biology*, 13(1), 99.

Wendte, J.M., Haag, J.R., Pontes, O.M., Singh, J., Metcalf, S. & Pikaard, C.S., 2019a. The Pol IV largest subunit CTD quantitatively affects siRNA levels guiding RNA-directed DNA methylation. *Nucleic Acids Research*, 47(17), 9024–9036.

Wendte, J.M., Haag, J.R., Singh, J., McKinlay, A., Pontes, O.M. & Pikaard, C.S., 2017. Functional Dissection of the Pol V Largest Subunit CTD in RNA-Directed DNA Methylation. *Cell Reports*, 19(13), 2796–2808.

Wendte, J.M., Zhang, Y., Ji, L., Shi, X., Hazarika, R.R., Shahryary, Y., Johannes, F. & Schmitz, R.J., 2019b. Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation D. Baulcombe & D. Weigel, eds. *eLife*, 8, e47891.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., et al., 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal*, 27(6), 581–590.

Wielopolska, A., Townley, H., Moore, I., Waterhouse, P. & Helliwell, C., 2005. A high-throughput inducible RNAi vector for plants. *Plant Biotechnology Journal*, 3(6), 583–590.

Wierzbicki, A.T., Blevins, T. & Swiezewski, S., 2021. Long Noncoding RNAs in Plants. *Annual Review of Plant Biology*.

Wierzbicki, A.T., Haag, J.R. & Pikaard, C.S., 2008. Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell*, 135(4), 635–648.

Williams, B.P., Pignatta, D., Henikoff, S. & Gehring, M., 2015. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. *PLOS Genetics*, 11(3), e1005142.

Wu, H., Li, B., Iwakawa, H., Pan, Y., Tang, X., Ling-hu, Q., Liu, Y., Sheng, S., Feng, L., Zhang, H., et al., 2020. Plant 22-nt siRNAs mediate translational repression and stress adaptation. *Nature*, 581(7806), 89–93.

Wu, L., Mao, L. & Qi, Y., 2012. Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. *Plant Physiology*, 160(2), 990–999.

Wu, Y.-Y., Hou, B.-H., Lee, W.-C., Lu, S.-H., Yang, C.-J., Vaucheret, H. & Chen, H.-M., 2017. DCL2- and RDR6-dependent transitive silencing of SMXL4 and SMXL5 in Arabidopsis dcl4 mutants causes defective phloem transport and carbohydrate over-accumulation. *The Plant Journal*, 90(6), 1064–1078.

Xia, R., Chen, C., Pokhrel, S., Ma, W., Huang, K., Patel, P., Wang, F., Xu, J., Liu, Z., Li, J., et al., 2019. 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nature Communications*, 10(1), 627.

Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A. & Ohad, N., 2019. RdDM - independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. *Nature Communications*, 10(1), 1–10.

Yan, H., Chretien, R., Ye, J. & Rommens, C.M., 2006. New Construct Approaches for Efficient

Gene Silencing in Plants. *Plant Physiology*, 141(4), 1508–1518.

Yan, X., Dong, X., Liu, L., Yang, Y., Lai, J. & Guo, Y., 2016. DNA methylation signature of intergenic region involves in nucleosome remodeler DDM1-mediated repression of aberrant gene transcriptional read-through. *Journal of Genetics and Genomics*, 43(8), 513–523.

Ye, R., Chen, Z., Lian, B., Rowley, M.J., Xia, N., Chai, J., Li, Y., He, X.-J., Wierzbicki, A.T. & Qi, Y., 2016. A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis. *Molecular Cell*, 61(2), 222–235.

Yoshikawa, M., Iki, T., Numa, H., Miyashita, K., Meshi, T. & Ishikawa, M., 2016. A Short Open Reading Frame Encompassing the MicroRNA173 Target Site Plays a Role in trans-Acting Small Interfering RNA Biogenesis. *Plant Physiology*, 171(1), 359–368.

Yoshikawa, M., Peragine, A., Park, M.Y. & Poethig, R.S., 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes & Development*, 19(18), 2164–2175.

Yuan, C., Wang, J., Harrison, A.P., Meng, X., Chen, D. & Chen, M., 2015. Genome-wide view of natural antisense transcripts in Arabidopsis thaliana. *DNA Research*, 22(3), 233–243.

Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. & Zilberman, D., 2013. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell*, 153(1), 193–205.

Zemach, A., McDaniel, I.E., Silva, P. & Zilberman, D., 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, 328(5980), 916–919.

Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T., Teng, C., Chen, X., Park, S.Y., Liu, L., Gallego-Bartolome, J., et al., 2015. A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell*, 163(2), 445–455.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J. & Jacobsen, S.E., 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proceedings of the National Academy of Sciences*, 104(11), 4536–4541.

Zhang, X., Xia, J., Lii, Y., Barrera-Figueroa, B., Zhou, X., Gao, S., Lu, L., Niu, D., Liang, W., Chen, Z., et al., 2012. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biology*, 13(3), R20.

Zhang, X., Zhu, Y., Liu, X., Hong, X., Xu, Y., Zhu, P., Shen, Y., Wu, H., Ji, Y., Wen, X., et al., 2015. Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in Arabidopsis. *Science*, 348(6230), 120–123.

Zhang, Y., Harris, C.J., Liu, Q., Liu, W., Ausin, I., Long, Y., Xiao, L., Feng, L., Chen, X., Xie, Y., et al., 2018. Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis. *Proceedings of the National Academy of Sciences*, 201716300.

Zhang, Z., Liu, X., Guo, X., Wang, X.-J. & Zhang, X., 2016. Arabidopsis AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing. *Nature Plants*, 2(5), nplants201649.

Zhao, M., San León, D., Delgadillo, Ma.O., García, J.A. & Simón-Mateo, C., 2014. Virus-induced gene silencing in transgenic plants: transgene silencing and reactivation associate with two patterns of transgene body methylation. *The Plant Journal*, 79(3), 440–452.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.-S. & Gregory, B.D., 2010. Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genetics*, 6(9).

Zhong, X., Du, J., Hale, C.J., Gallego-Bartolome, J., Feng, S., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J. & Jacobsen, S.E., 2014. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. *Cell*, 157(5), 1050–1060.

Zhou, M., Palanca, A.M.S. & Law, J.A., 2018. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nature Genetics*, 1.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.-H., Liou, L.W., Barefoot, A., Dickman, M. & Zhang, X., 2011. Arabidopsis Argonaute10 Specifically Sequesters miR166/165 to Regulate Shoot Apical Meristem Development. *Cell*, 145(2), 242–256.

Zuo, J., Niu, Q. & Chua, N., 2000. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant Journal*, 24(2), 265–273.

Attachments

Čermák and Fischer *BMC Plant Biology* (2018) 18:252
https://doi.org/10.1186/s12870-018-1482-3

BMC Plant Biology

RESEARCH ARTICLE Open Access

Pervasive read-through transcription of *T-DNAs* is frequent in tobacco BY-2 cells and can effectively induce silencing

Vojtěch Čermák and Lukáš Fischer*

Abstract

Background: Plant transformation via *Agrobacterium tumefaciens* is characterized by integration of commonly low number of *T-DNAs* at random positions in the genome. When integrated into an active gene region, promoterless reporter genes placed near the *T-DNA* border sequence are frequently transcribed and even translated to reporter proteins, which is the principle of promoter- and gene-trap lines.

Results: Here we show that even internal promotorless regions of *T-DNAs* are often transcribed. Such spontaneous transcription was observed in the majority of independently transformed tobacco BY-2 lines (over 65%) and it could effectively induce silencing if an inverted repeat was present within the *T-DNA*. We documented that the transcription often occurred in both directions. It was not directly connected with any regulatory elements present within the *T-DNAs* and at least some of the transcripts were initiated outside of the *T-DNA*. The likeliness of this read-through transcription seemed to increase in lines with higher *T-DNA* copy number. Splicing and presence of a polyA tail in the transcripts indicated involvement of Pol II, but surprisingly, the transcription was able to run across two transcription terminators present within the *T-DNA*. Such pervasive transcription was observed with three different *T-DNAs* in BY-2 cells and with lower frequency was also detected in *Arabidopsis thaliana*.

Conclusions: Our results demonstrate unexpected pervasive read-through transcription of *T-DNAs*. We hypothesize that it was connected with a specific chromatin state of newly integrated DNA, possibly affected by the adjacent genomic region. Although this phenomenon can be easily overlooked, it can have significant consequences when working with highly sensitive systems like RNAi induction using an inverted repeat construct, so it should be generally considered when interpreting results obtained with the transgenic technology.

Keywords: GFP, Inverted repeat, Promoterless, RNAi, Read-through transcription, *T-DNA*, Tobacco BY-2 cell line

Background

Agrobacterium tumefaciens mediated transformation is a common method used to obtain transgenic plants. *Agrobacterium* transfers its *T-DNA* into a plant cell, where it can be integrated inside plant genome predominantly through double-strand break repair pathway (reviewed in [1]). Generally *T-DNAs* are stably introduced only in a small proportion of cells cocultivated with *Agrobacterium*. The *T-DNAs* are integrated at random positions in their genome [2, 3]. Subsequent regeneration of transgenic lines requires incorporation of a selection step to filter out untransformed cells/plants. This selection is usually achieved by incorporating antibiotic or herbicide resistance gene into the *T-DNA*. The requirement for the selection gene to be actively expressed then imposes bias on the selected transformants. *T-DNAs* of such transformants are preferentially present in regions with active transcription, especially near promoters and in regions with low nucleosome density [3, 4]. This probably leads to the unusually high success rate of various promoter- and gene-trap lines [5–7]. Although the transformants generated by *Agrobacterium* have lower number of insertions compared to other transformation methods, there are still many transformants with multiple insertions. Commonly the number of insertions per line varies between 1.4 and 4.9 [4, 8]. In case of multiple insertions, it is quite common for *T-DNAs* to integrate in a form of direct or inverted

* Correspondence: lukas.fischer@natur.cuni.cz
Department of Experimental Plant Biology, Charles University, Faculty of Science, Viničná 5, 128 44 Prague 2, Czech Republic

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

repeats in one position in the genome. Due to the way how *T-DNAs* are integrated, the most common form is the head-to-head (RB-to-RB) inverted repeat arrangement [9, 10]. Convergent read-through transcription of *T-DNAs* integrated as inverted repeats can induce silencing of homologous sequences via RNA interference (RNAi) [11, 12].

RNAi is an important mechanism in the regulation of gene expression in eukaryotic cells. In functional genomics, it is often used as a tool to modify expression of studied genes. The key players in RNAi are small RNAs (sRNAs), which can be formed by multiple pathways in plants, making the plant RNAi a very complex process (reviewed in [13]). Generally, a double stranded RNA (dsRNA) is needed to induce sRNA production in plants. There are many different ways to achieve dsRNA formation in a cell, the most efficient one is intermolecular pairing of transcripts coming from an inverted repeat [14].

Triggering RNAi by introduced "silencer constructs" can be used to knock out genes of interest or to study the mechanisms of RNAi itself. The large majority of RNAi studies were based on the model plant *Arabidopsis thaliana* that offers high quality genomic data and plenty of mutant lines, which are easily accessible to the research community. Few years ago, we started to test an alternative model, tobacco BY-2 cell line that has been successfully used in numerous studies focused on cellular processes [15]. The BY-2 cell culture is composed of relatively homogeneous mitotically proliferating cells [15, 16]. The absence of the gametophytic phase also prevents some types of epigenetic changes connected with this developmental stage [17–20]. The BY-2 cell line also allows simple analyses at the level of individual cells and assessment of a large number of independent transgenic lines (in the callus form) that can be easily generated, managed and analyzed [16]. Since the behavior of individual transgenic lines of any model is affected by the *T-DNA* copy number and the chromosomal environment of the insertion [21–23], analysis of a high number of independent transformed lines, which provides a more generalized picture, is recommended. Other advantages include easy and reliable analyses of fluorescence levels in these cells as well as simple ways to treat these cells with various chemicals. In situations, where the study is focused on analyses of general molecular and cellular mechanisms, the absence of the whole plant context may not have substantial impact on the appropriate generalization of the results.

The observation of pervasive read-through transcription of *T-DNAs* that we describe in this study, was discovered during our RNAi project focused on comparison of silencing potential of different silencing inducers. We super-transformed a BY-2 line stably expressing the *GFP* gene [24] with various silencers that were not controlled by a constitutive promoter as usual, but they were based on the XVE inducible system [25, 26], providing the possibility of highly reliable induction of RNAi by β-estradiol. The *GFP* reporter gene was used to allow simple visualization of silencing.

We generated hundreds of independent transgenic BY-2 lines (calli) with a goal to assess population responses to the induced expression of each silencing inducer (these data are not presented in this study). Surprisingly, we observed significant differences between the silencers already prior to their activation with β-estradiol with high frequency of silencing occurring in the calli carrying the inverted repeat construct. We found that this silencing correlated with spontaneous transcription of the silencer and here we show detailed analysis of this phenomenon.

Methods

Plasmid construction

All silencer *T-DNAs* (Fig. 1) were prepared similarly. First the full length *GFP* sequence was PCR amplified from *psmRS-GFP* plasmid [27] using primers with appropriate adapters (Additional file 1). The sequence was then inserted into the *pDrive* vector (part of QIAGEN PCR Cloning Kit) and subsequently transferred into the destination binary vector – either *pER8* [26] or *pGreen* [28] using the appropriate restriction endonucleases. The *AS-GFP* was cloned between *Xho*I and *Bcu*I in *pER8*, the *UT-GFP* was cloned between *Xho*I and *Pvu*II in *pER8*, the *GFP* was cloned between *Xho*I and *Bcu*I in *pER8* (as *Sal*I and *Nhe*I fragment derived from *pDrive IR-GFP*). The *IR-GFP* construct was first assembled in *pDrive*: the intron from *Solanum tuberosum PsbO* gene (cDNA GeneBank no. X17578.1) including approximately 20 bp from exons on both sides was AT-cloned into *pDrive* (in *Sac*I-*Kpn*I orientation). The intron sequence placed inside the inverted repeats was demonstrated to enhance the efficiency of silencing likely via facilitating dsRNA formation [29]. *IR1-GFP* was inserted first between *Xho*I and *Sal*I, then *IR2-GFP* was inserted between *Pst*I and *Bam*HI; the whole *IR-GFP* construct was then cloned between *Xho*I and *Bcu*I in *pER8*. The *IR-GFP-ΔP1* control was created by cloning the *IR-GFP* construct between *Sal*I and *Bcu*I in *pER8* (*Sal*I cleaves at the beginning of the inducible promoter). The *IR-GFP-ΔP2* control was created by cloning the *IR-GFP* construct in *pGreen 0129* between *Sac*I and *Kpn*I. All cloning experiments were completed using enzymes from Fermentas (Thermo Fisher Scientific) and *Escherichia coli* strain JM109. The resulting *T-DNAs* were confirmed by restriction and sequencing (Fig. 1).

Plant material

Tobacco cell line BY-2 (*Nicotiana tabacum* L. cv. Bright Yellow 2) [15] was obtained from Prof. Zdeněk Opatrný, who had cultivated the line for more than 20 years. BY-2 calli were cultivated on agar plates (0.8 *w*/*v* agar; 6 cm

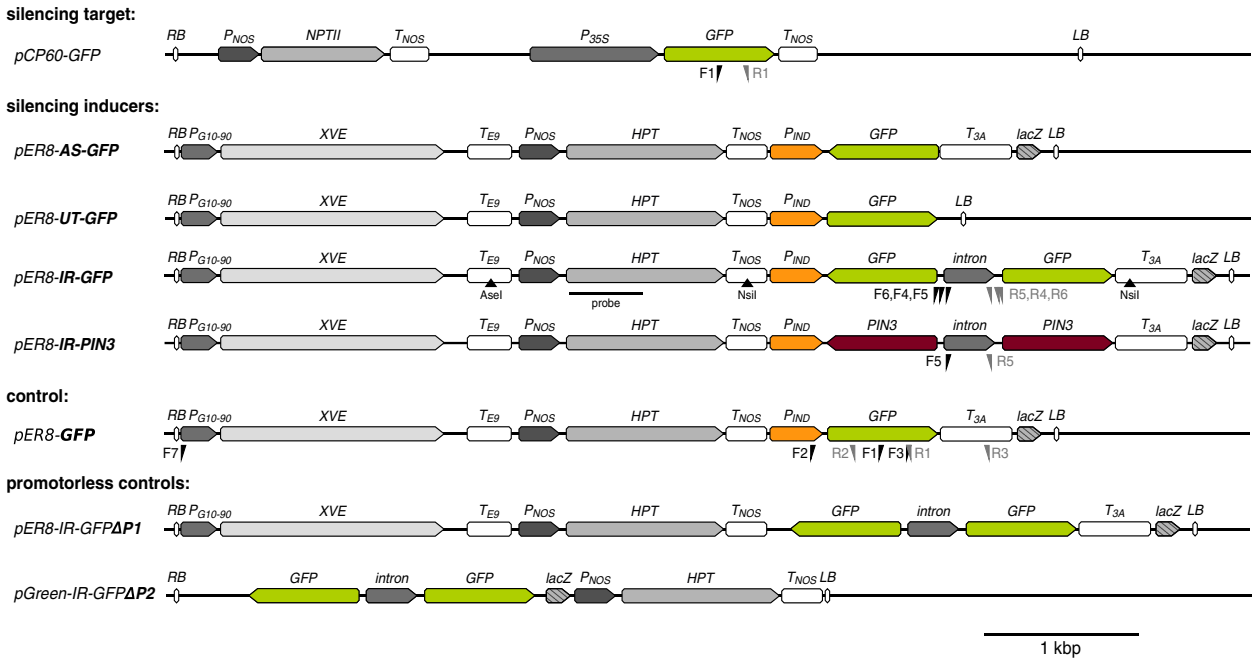


Fig. 1 Schematic representation of *T-DNAs* used in this study. Functional elements, positions of primers used for reverse transcription and qPCR, position of a probe and restriction enzyme sites used for Southern hybridization are shown above and below the appropriate sequences. *pCP60-GFP*: construct for expression of *GFP* under the constitutive promoter. *pER8*: a set of *T-DNAs* used for the inducible expression: *pER8-AS-GFP*: sequence of *GFP* in antisense orientation; *pER8-UT-GFP*: sequence of *GFP* without terminator; *pER8-IR-GFP*: sequence of *GFP* arranged as inverted repeat; *pER8-IR-PIN3*: sequence of *PIN3* gene arranged as inverted repeat; *pER8-GFP*: sense *GFP*. The *pER8-IR-GFP*ΔP1 and *pGreen-IR-GFP*ΔP2 are the two promoterless controls with *GFP* sequence arranged as inverted repeat. *RB*: right border, *LB*: left border, P_{NOS}: nopaline synthase promoter, P_{35S}: *35S* promoter, $P_{G10–90}$: synthetic constitutive promoter; P_{IND}: inducible promoter activated by β-estradiol; T_{NOS}: nopaline synthase terminator; T_{E9}: *rbcS E9* terminator; T_{3A}: *rbcs S 3A* terminator; *XVE*: estrogen receptor and transcriptional activator; *NPTII*: kanamycin resistance gene; *HPT*: hygromycin resistance gene; *GFP*: green fluorescent protein coding sequence; *intron*: intron from *Solanum tuberosum PsbO* gene; *LacZ*: fragment of bacterial β–galactosidase gene. All *T-DNAs* are at the same scale as indicated with 1kbp scale bar

diameter plates) with modified MS medium [30]. The calli were subcultured monthly. Suspension cell cultures were subcultured every seventh day (1 ml of cells into 30 ml of liquid media). The cultures were kept in darkness at 26 °C; suspensions were placed on the orbital shaker IKA KS501 at 110 rpm (IKA Labortechnik, Staufen, Germany; orbital diameter 30 mm).

Transformations of BY-2 suspension cells were carried out as described previously [24] using *Agrobacterium tumefaciens* strain C58C1 carrying a helper plasmid *pGV2260* [31] and appropriate binary vector (see above). After cocultivation with agrobacterium, the cells were plated on solidified medium containing 25 μg/ml hygromycin and 100 μg/ml cefotaxim and cultured for 3 weeks. Using this procedure, individual transformed cells form isolated macroscopic cell clusters (commonly called calli) that can be mostly regarded as genetically homogeneous clones [24].

The promoter from the *pER8* plasmid was induced by cultivating the calli on media with addition of 2 μM β-estradiol (from Sigma-Aldrich Cat. No. E2758). The β-estradiol was stored as 20 mM solution in DMSO, therefore, a corresponding amount of DMSO was added to the cultivation medium of the controls.

Arabidopsis thaliana (Col-0) plants were grown in Jiffy soil pellets under long-day conditions with illumination of 100 μm m^{-2} s^{-1} photosynthetically active radiation (OSRAM L 58 W/930).

Measurement and analysis of fluorescence level

The BY-2 calli grown after transformation (see above) were transferred to new plates with 100 μg/ml cefotaxim (20 calli per plate) and cultivated for approximately 8 days. 200 calli (each representing different transgenic event) were used for each of the variants. Each plate was photo-documented separately using G:BOX (SynGene, Cambridge, UK) with blue excitation light (LED diodes with maximum at λ = 465 nm) and green emission filter (FILT525/GX; 510–540 nm). The images were processed using software NIS-Elements 3.10 (Laboratory Imaging, Prague, CZ). The average light intensity was measured for all the pixels from each callus. These data were statistically analyzed using R 3.1.2 and Pearson's chi-square test. The threshold for fluorescent and non-florescent calli was set as the highest fluorescence intensity measured for wild-type BY-2 callus (these were used as controls cultivated alongside the transgenic calli).

Transcription analysis

RNA was isolated from 100 mg of 10 to 18 days old calli (in the same experiment the calli were always in the same age) or from 50 mg of siliques of 40 days old *Arabidopsis thaliana* plants (younger siliques were selected – less than 10 days old, to ensure high proportion of dividing cells in the tissue) using NucleoSpin® RNA Plant kit (MACHEREY-NAGEL, Düren, DE). The procedure was performed according to manufacturer's instructions including the on-column DNA digestion. The RNA was measured on NanoDrop 2000 (Thermo Fisher Scientific) to assess the concentration of the samples and to exclude contamination of the RNA by impurities. The integrity of the RNA was checked by gel electrophoresis using the "bleach gel" method [32]. For cDNA preparation, 1 μg of the total RNA was again treated with DNase I and half of the reaction mixture was then used as template for RevertAid Reverse Transcriptase (Fermentas, Thermo Fisher Scientific). The other half was exposed to the same treatment, but without adding the Reverse Transcriptase – this served as a control to check for DNA contamination. The final cDNA was diluted into the volume of 50 μl. Either oligo(dT) or specific primers were used for the cDNA synthesis. Some of the specific primers were designed to allow for distinction between sense and antisense transcripts and spliced and unspliced molecules (Additional file 1).

The quantification itself was done by qPCR, using LightCycler 480 (Roche) and iQ™ SYBR® Green Supermix (BioRad, Hercules, CA, USA). All the experiments were done while keeping the general qPCR guidelines in mind [33]. Reactions were completed in 10 μl volume, using 1 μl of cDNA as a template; all reactions were done in triplicate. The specificity of the PCR was verified by melting curve analysis (using the LightCycler 480 software) and also by checking randomly selected samples using gel electrophoresis. For each set of primers, there was appropriate negative control (WT and/or dH_2O). The PCR efficiency for each amplicon and the Cq values for each sample were calculated using the software LinRegPCR 2015.3 [34]. The values for triplicates were averaged after correction for PCR efficiency. Samples in the triplicate with no amplification or only unspecific products were counted as zeroes; the samples with the majority of unspecific product was treated as one order of magnitude smaller (unspecific products appeared only for samples with high Cq values, over 30). Calculated concentrations were normalized to the expression of *NtEF1α*, so all the presented values show the relative level of given transcript to the level of *NtEF1α*. Results were then statistically compared using R 3.1.2 and Welch's t-test. Positions of primers used for qPCR are indicated in Fig. 1 and their sequences listed in Additional file 1. Some primer sequences were taken over from previous studies [35, 36].

The BY-2 calli for transcriptional analysis were selected randomly from groups of silenced and non-silenced calli based on the presence or absence of GFP fluorescence. In case of transformants carrying sense *GFP* (*GFP* sequence in *pER8* XVE inducible system), the BY-2 calli and *Arabidopsis* lines were randomly selected from those that were able to induce *GFP* expression when grown on the induction medium supplemented with β-estradiol.

Southern blot analysis

The Southern blot hybridization was done as described previously [37] with the following modifications: The DNA was isolated from 150 mg (FW) of BY-2 calli. 20 μg of genomic DNA per sample was separately digested by enzymes *Nsi*I and *Ase*I (New England Biolabs). The probe was prepared as a fragment of HPT gene using PCR with primers *HPT_probe_F* and *HPT_probe_R* (Additional file 1).

The Southern blot was interpreted as follows: tandem *T-DNA* inserted as direct repeat should give 6.8 kbp fragment with both *Nsi*I and *Ase*I, plus one fragment of unknown size for *Nsi*I and *Ase*I; head-to-head inverted repeat should give 7.3 kbp fragment when digested with *Nsi*I and two fragments of unknown size when digested with *Ase*I; tail-to-tail inverted repeat should give 9.6 kbp fragment when digested with *Ase*I and two fragments of unknown size when digested with *Nsi*I.

Results

Fluorescence in calli transformed with various *GFP* silencer constructs

To study various aspects of RNAi, we prepared three different silencer *T-DNAs* based on the XVE inducible system [26]. Specifically, the silencing should have been achieved through production of i) antisense RNA (*AS-GFP*), ii) non-polyadenylated sense RNA (*GFP* without any terminator; *UT-GFP*) and iii) hairpin RNA (inverted repeat with an intron separating the antisense and sense *GFP* fragment; *IR-GFP*). These *T-DNAs* were expected to trigger posttranscriptional silencing of the reporter *GFP* gene only under induction with β-estradiol. As a control, we also prepared a construct with inducible *GFP* in sense orientation and ended with terminator (*GFP*; Fig. 1).

A selected BY-2 cell line that has been stably expressing *GFP* (driven by *35S* promoter) for more than 8 years [24] was separately supertransformed with each *T-DNA*. We then analyzed GFP fluorescence in individual calli grown after the transformation - each representing independent transformation event (see Methods for details). On the control medium, where the calli were not exposed to β-estradiol, we expected similar fluorescence in all populations (hundreds of calli) carrying various silencers. However, we observed that *IR-GFP* population had strikingly

lower frequency of calli with detectable GFP fluorescence compared to the other silencer and control *T-DNAs* (Fig. 2). All the differences between frequencies of GFP-positive calli in the IR-GFP population and those with other constructs were statistically significant ($p < 10^{-40}$). It should also be noted that some of the smaller differences between the variants were found significant as well: the UT-GFP variant compared to EV and GFP ($p < 10^{-6}$).

Since the calli grew on the medium without β-estradiol, the observed unexpectedly high proportion (over 65%) of spontaneously silenced IR-GFP calli prompted us to study this phenomenon further. Although XVE inducible system is considered to be reliable with very low leakiness of the inducible promoter [26], we had to exclude this possibility. We prepared two additional controls; i) we removed the inducible promoter from the *pER8* vector (*IR-GFPΔP1*) and ii) we cloned the promoterless *IR-GFP* into the empty *pGreen* vector (*IR-GFPΔP2*). After transformation to BY-2 cells, the number of spontaneously silenced independent calli was virtually the same as with the original *IR-GFP T-DNA* (Fig. 2, columns 5 and 6; the differences between the IR variants were not statistically significant), indicating that the silencing was independent of the presence of the inducible promoter and the *T-DNA* context.

To exclude the possibility that the transcription was initiated from elements that might be common for both *T-DNAs*, we compared their sequences. We found two homologous regions: i) the HPT expression cassette, which, however, differed between the two *T-DNAs* in its orientation relative to the *IR-GFP* sequence (Fig. 1) and ii) a short 156-nt fragment of bacterial β–galactosidase gene (*LacZ*) that was downstream of the *IR-GFP* in both *T-DNAs*. No promoter regulatory elements were predicted within this sequence by TSSP software (http://linux1.softberry.com).

Transcription and splicing of *IR-GFP*

Transcription analysis was done in five non-silenced and five silenced calli that were randomly selected from populations with and without detectable GFP fluorescence. The results showed that the levels of *GFP* transcripts roughly matched the GFP fluorescence intensities, with the lowest transcription being detected in the silenced calli (Fig. 3a and b). To see whether the silencing correlated with transcription of the *IR-GFP*, we generated cDNAs using primers specific for the *GFP* hairpin transcribed in both the sense and antisense orientation (the "sense" and "antisense" transcripts were relative to the intron separating the *GFP* sequences in the *IR-GFP*). Transcript levels were analyzed using qPCR (Fig. 3b-h) with primers designed to allow separate quantification of i) spliced transcripts, ii) unspliced transcripts and iii) intron-containing molecules (i.e. nascent transcripts and spliced introns; Fig. 1). The *IR-GFP* transcripts were detected in all the spontaneously silenced calli at levels even higher than in the callus, where the *IR-GFP* transcription was induced with β-estradiol. In contrast, almost undetectable levels (three to four orders of magnitude lower) were found in all non-silenced calli with detectable GFP fluorescence (Fig. 3a and c, for the comparison of averaged relative transcript levels and their statistical comparison see Additional file 2). Surprisingly, in three of the five silenced calli, the *IR-GFP* was clearly transcribed also in the "antisense" direction (from the terminator towards the promoter), although at lower levels than it was

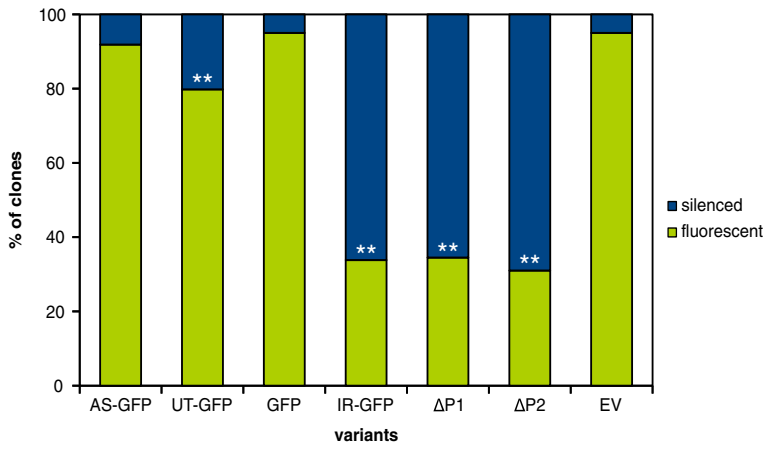


Fig. 2 Frequency of spontaneous silencing of *GFP* in supertransformed BY-2 calli. Silencer and control *T-DNAs* (as described in Fig. 1) were supertransformed to BY-2 cell line stably expressing *GFP*. The fluorescence was assessed in independently transformed calli grown on non-inductive medium six weeks after transformation. The bars represent the percentage of fluorescent non-silenced calli (the lower part of the bar) and silenced calli (the upper part of the bar). The variants AS-GFP and UT-GFP represent averages of three biological replicas and the variants GFP and IR-GFP represent averages of four biological replicas (independent transformations). Each replica had 200 calli (with exception of EV with 160 calli). Variants that significantly differed from the EV control are marked with ** ($p < 10^{-6}$). ΔP1: pER8-*IR-GFP*ΔP1; ΔP2: pGreen-*IR-GFP*ΔP2; EV: empty vector *pER8*

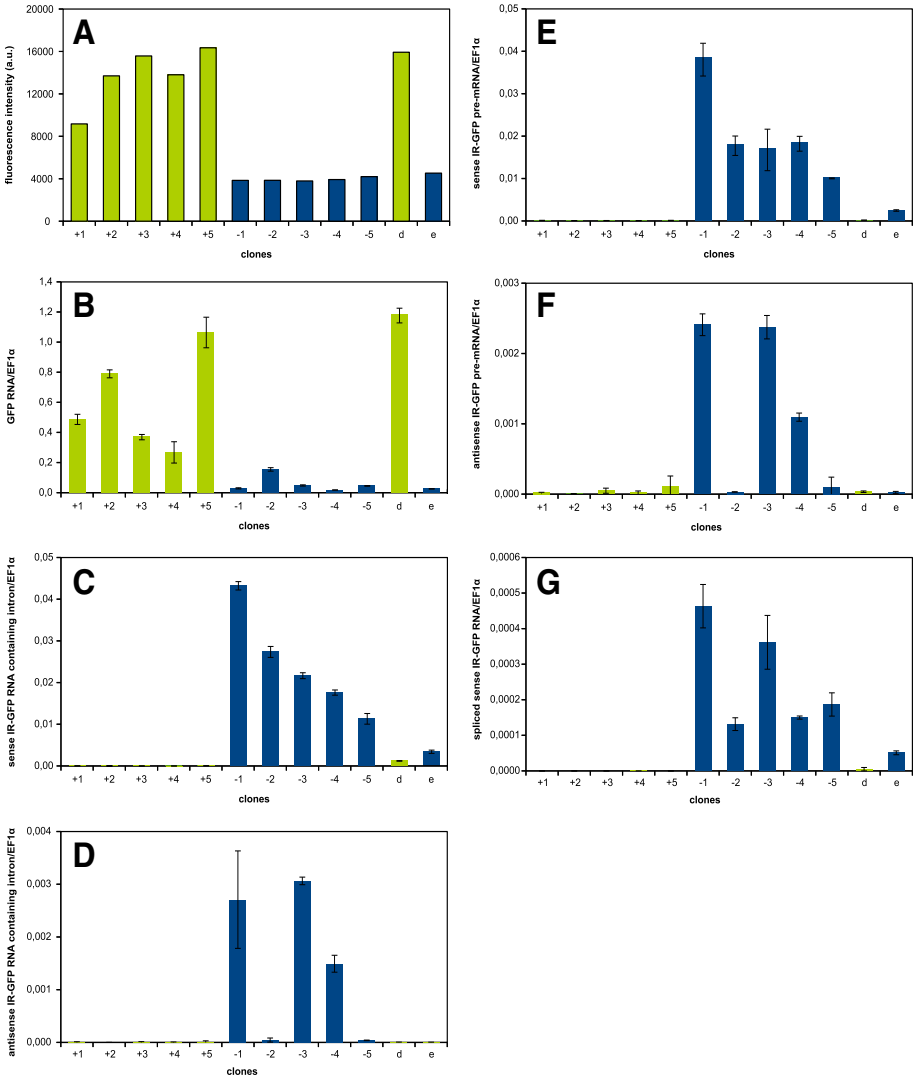


Fig. 3 Analysis of selected BY-2 calli supertransformed with *pER8-IR-GFP*. RT-qPCR analysis of *IR-GFP* transcript levels in five randomly selected spontaneously silenced calli (marked from − 1 to − 5) and five non-silenced calli (marked from + 1 to + 5) and one additional non-silenced callus grown on the induction medium with β-estradiol (marked as "e") and on the control medium with DMSO (marked as "d"). Each callus represented independent transformation event. **a** Fluorescence of analyzed calli (in arbitrary units), signals below 4000 represent the background; **b** the level of the sense *GFP* transcript (RT: oligo dT primer, qPCR: F1 and R1 primers); **c** the level of transcripts with the sense intron of *IR-GFP* (representing both the unspliced transcript and the spliced intron; RT: R5 primer, qPCR: F5 and R5 primers); **d** the level of transcripts with the antisense intron of *IR-GFP* (RT: F5 primer, qPCR: F5 and R5 primers); **e** the level of the sense *IR-GFP* unspliced transcript (RT: R5 primer, qPCR: F4 and R5 primers); **f** the level of the antisense *IR-GFP* unspliced transcript (RT: F5 primer, qPCR: F5 and R4 primers); **g** the level of the sense spliced transcript (RT: R6 primer, qPCR: F6 and R6 primers)

transcribed in the "sense" orientation. There was no impact of β-estradiol treatment on the hairpin transcription in the antisense orientation in the control line (Fig. 3d).

The results obtained from detecting the intron sequence were similar to those obtained from detecting the unspliced transcript (PCR product spanning the exon-intron boundary), this was also true for the antisense direction (Fig. 3e and f). Such results could indicate either fast degradation of the spliced intron or that the intron was not spliced at all. To see if the splicing actually took place, we ran qPCR over the supposed exon-exon boundary. Using the "sense" cDNA, we were indeed able to amplify the specific product corresponding to the spliced hairpin RNA in the silenced calli at similar levels in both the spontaneously silenced calli and in the estradiol-induced control. The ratio of spliced/unspliced transcripts was between 0.7 and 2.1% for the spontaneously transcribed lines and 2.1% for the estradiol-induced control. It suggested that at least some of the *IR-GFP* transcripts were spliced correctly in the spontaneously silenced calli (Fig. 3g; the specificity of the qPCR was verified, see Methods).

Spontaneous transcription of *IR-PIN3* and sense *GFP*

The spontaneous transcription and subsequent silencing could be specific for our *IR-GFP* construct or it could represent more general phenomenon. Therefore, we tested the occurrence of spontaneous transcription of another inverted repeat construct, *IR-PIN3* prepared from a fragment of tobacco *PIN3* gene placed in the *pER8* plasmid (*NtPIN3bT*; kindly provided by Jan Petrášek) that we used to transform wild-type BY-2 cells. We analyzed the expression of the native *PIN3* gene and the *PIN3* hairpin in 11 randomly selected independent calli using RT-qPCR. We observed expression of the *PIN3* hairpin in the majority of analyzed calli. The relative transcription levels seemed to be somewhat lower (from 2 to 20 times lower) compared to the *GFP* hairpin in the *IR-GFP* silenced calli (Fig. 4a-c, for the comparison of averaged relative transcript levels see Additional file 2). However, the comparison was based on the amplification of the intron sequence, so the real transcription levels of the two hairpins could differ due to various efficacy of their splicing. The *IR-PIN3* transcription did not cause strong silencing of *PIN3* in contrast to the situation with *IR-GFP* and *GFP*. Only a weak decrease in *PIN3* expression could be observed in some calli with the highest level of the PIN3 hairpin (see Fig. 4a and b). However, the averaged *IR-PIN3* transcript levels did not significantly differ between calli with higher and lower *PIN3* transcription (Additional file 2).

To assess if the observed high frequency of spontaneous transcription was specifically connected with inverted repeat arrangement of introduced transgenes or if it was more general phenomenon in our experimental system, we analyzed *T-DNA* transcription in lines carrying sense *GFP* in *pER8* XVE inducible system (Fig. 1). After transformation into wild-type BY-2 cell line, there were no calli with detectable GFP fluorescence that would indicate spontaneous transcription connected with subsequent translation into functional GFP protein without β-estradiol treatment. However, RT-qPCR analysis showed that, at the transcriptional level, the sense *GFP* construct behaved similarly to the inverted repeats. The transcription was detected in all five calli and in both sense and antisense directions. The relative transcript levels were similar to those described for the

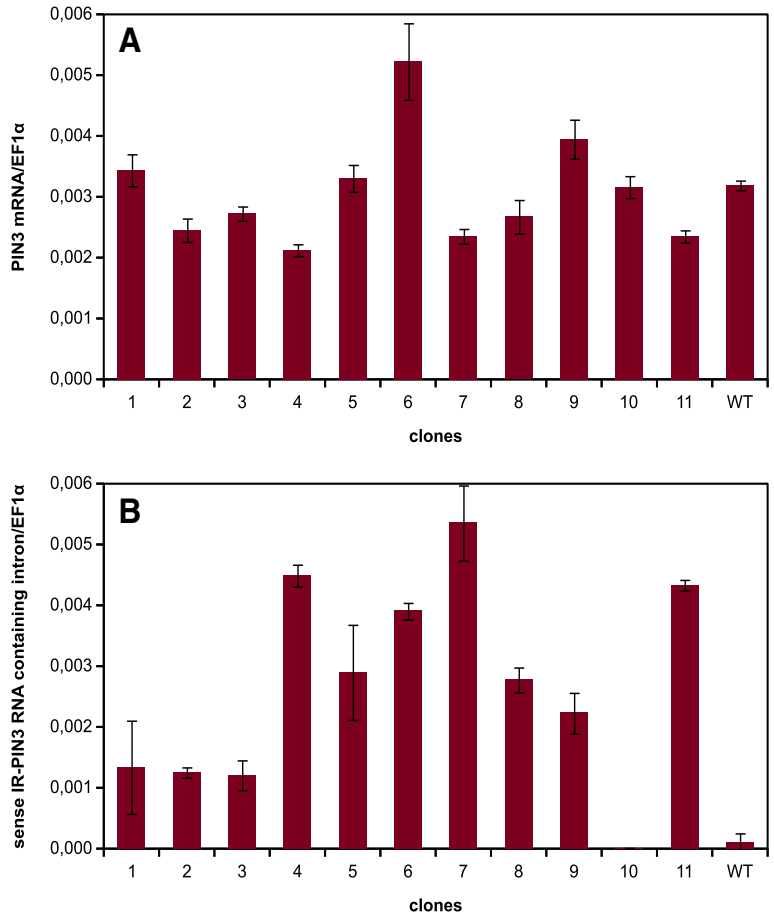


Fig. 4 Transcription analysis of selected BY-2 calli transformed with *pER8-IR-PIN3*. RT-qPCR analysis of transcript levels in eleven randomly selected calli and an untransformed BY-2 callus as a control. **a** The level of the *PIN3* mRNA (RT: oligo dT primer, qPCR: PIN3_F and PIN3_R primers); **b** the level of the sense intron of the IR-PIN3 (representing both the unspliced transcript and the spliced intron; RT: R5 primer, qPCR: F5 and R5 primers)

hairpins (Fig. 3c-f, Fig. 5a and c, Additional file 2). This showed that the transcription was not connected with the inverted-repeat character of the sequence present in the *T-DNA*.

To better understand the nature of the spontaneous transcription, we tried to identify the transcription start site by 5'RACE, but the attempts failed despite intensive optimization. Therefore, we investigated whether the transcript originated within or outside the transcription unit in *pER8 T-DNA*. For this purpose, we used a primer that matched to the region 50 nt upstream of the transcription start site (*TSS*) for the sense transcript and a primer matching to the *3A* terminator region 50 nt downstream of the last predicted poly(A) signal for the antisense transcript (as predicted by PASPA software: http://bmi.xmu.edu.cn/paspa/index.html; [38]). We detected transcripts from both regions and their levels in individual calli correlated with previously detected transcripts of the *GFP* gene. This suggested that the spontaneous transcription originated (at least partially) outside of the transcription unit in *pER8 T-DNA*. The nearest *ATG* is 99 bp upstream from the proper *ATG*, so transcripts originating from this 99-bp long region could be theoretically translated into the proper GFP protein, but we have never detected GFP fluorescence in such lines, which indicated *TSS* being more upstream. Thus, we tried to roughly localize the position of *TSS* within the *T-DNA*. We designed a set of forward primers along the *T-DNA* for amplification from *cDNA* prepared with a reverse primer specific to the *GFP* sequence. Surprisingly, we obtained products even with the most upstream primer located near the border sequence and preceding any promoter present in the *T-DNA* (Fig. 6). This PCR product

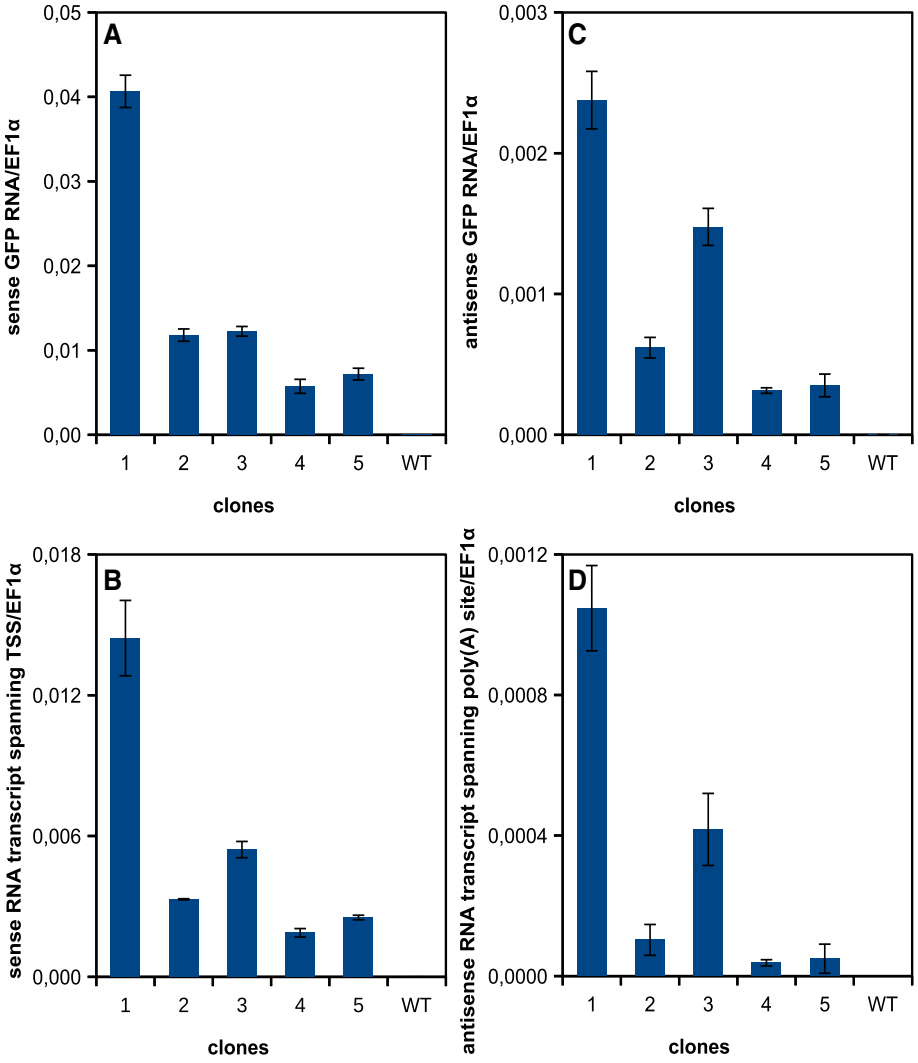


Fig. 5 Transcription of *GFP* gene in selected BY-2 calli transformed with *pER8-GFP*. RT-qPCR analysis of transcript levels in five selected independent calli grown on the medium without β-estradiol. **a** The level of the sense *GFP* transcript (RT: R1 primer, qPCR: F1 and R1 primers); **b** the level of the sense transcript containing the region 50 nt upstream of transcription start site (*TSS*) of the inducible promoter (RT: R1 primer, qPCR: F2 and R2 primers); **c** the level of the antisense *GFP* transcript (RT: F1 primer, qPCR: F1 and R1 primers); **d** the level of the antisense transcript containing the region 50 nt downstream of the last poly(A) signal of T3A terminator (RT: F1 primer, qPCR: F3 and R3 primers)

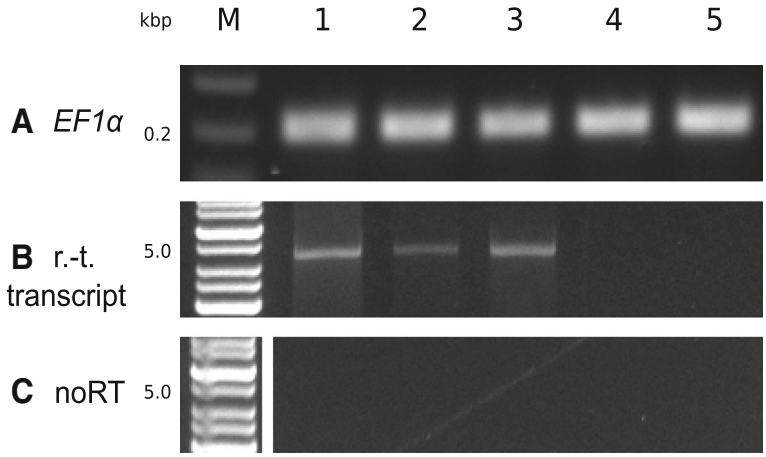


Fig. 6 Read-throught transcripts of the *pER8 T-DNA* in selected BY-2 calli transformed with *pER8-GFP*. Semiquantitative RT-PCR analysis of transcript levels in five independent calli untreated with β-estradiol (the same calli as in Fig. 5). cDNA was prepared using R1 primer. **a** The level of the *EF1α* transcript (internal standard); **b** the level of read-through (r.-t.) *pER8 T-DNA* transcript (F7 and R2 primers); **c** amplification of RNA samples that were not treated with reverse transcriptase to ensure that there was no DNA contamination

was more than 4 kbp long and included two upstream transcription units for the *XVE* receptor and the *HPT* gene. Thus the transcription was able to overcome two transcription terminators (Fig. 1).

Determination of *T-DNA* copy number and arrangement

To better understand potential reasons of this transcription, we estimated copy number and arrangement of *T-DNAs* by Southern hybridization in three silenced and three non-silenced lines (selected according to the data in Fig. 3). The results indicated that the read-through transcription and spontaneous silencing might be connected with higher *T-DNA* copy number (Additional file 3); 1 to 3 copies were detected in non-silenced lines and 3 to approximately 8 copies in silenced lines. Further analyzes of the size and number of hybridizing bands in individual lines suggested that *T-DNA* arrangements allowing read-through transcription from one *T-DNA* to another, i.e. direct *T-DNA* repeats and inverted tail-to-tail repeats, were not present in lines − 1 and + 2. In the other four lines the situation was not clear due to multiple insertions (− 3, − 4, + 3) or the presence of truncated copies (− 4, + 1), but theoretically direct repeat could be present in lines − 3 and − 4 and tail-to-tail repeat in line − 4 and + 3.

Spontaneous transcription in *Arabidopsis thaliana*

To assess wider significance of our observation we analyzed the transcription in a different model organism – *Arabidopsis thaliana* plants carrying the same *T-DNA* with sense *GFP* in *pER8* XVE inducible system. We randomly selected five transformants as before. RNA was isolated from immature siliques of plants grown in soil without any exposure to β-estradiol. As in the experiment with BY-2 cell line, we analyzed the *GFP* transcription in both the sense and antisense directions and also the transcripts from the regions spanning the canonical transcription start site and the poly(A) signal. We were able to detect transcripts in the sense direction from both the *GFP* sequence and from the region spanning the *TSS* and in the antisense direction from the *GFP* region (Fig. 7a-c). However transcription was detected in only one transformed line at the level comparable to the BY-2 callus with the lowest transcript level. The transcripts in the antisense direction over the poly(A) signal were almost undetectable (Fig. 7d). Transcription in leaves was somewhat lower (Additional file 4) than in immature siliques, which have higher proportion of actively dividing cells (similarly to the BY-2 cell line).

Discussion

GFP silencing was connected with spontaneous transcription of *IR-GFP T-DNA*

This study was prompted by observation of massive *GFP* silencing occurring after the introduction of *IR-GFP* construct into BY-2 cells with stable expression of *GFP*. Such observation was surprising because the *IR-GFP* was controlled by *XVE* system that is considered to be one of the most reliable, i.e. the least suffering from leaky transcription [25, 26]. Although it is rare in the XVE system, there are reports showing that leaky expression can occur [39, 40]. To exclude this possibility, we employed additional controls; two *IR-GFP* constructs without the presence of the inducible (or any other) promoter. The results were identical in both cases (Fig. 2), clearly showing that the silencing did not occur as a result of the inducible promoter leakiness.

The observed silencing of the *GFP* clearly correlated with transcription of the *IR-GFP*, suggesting that silencing occurred as a result of this transcription at the post-transcriptional level (Fig. 3). By testing *PIN3* inverted repeat, we excluded the possibility that the *GFP*

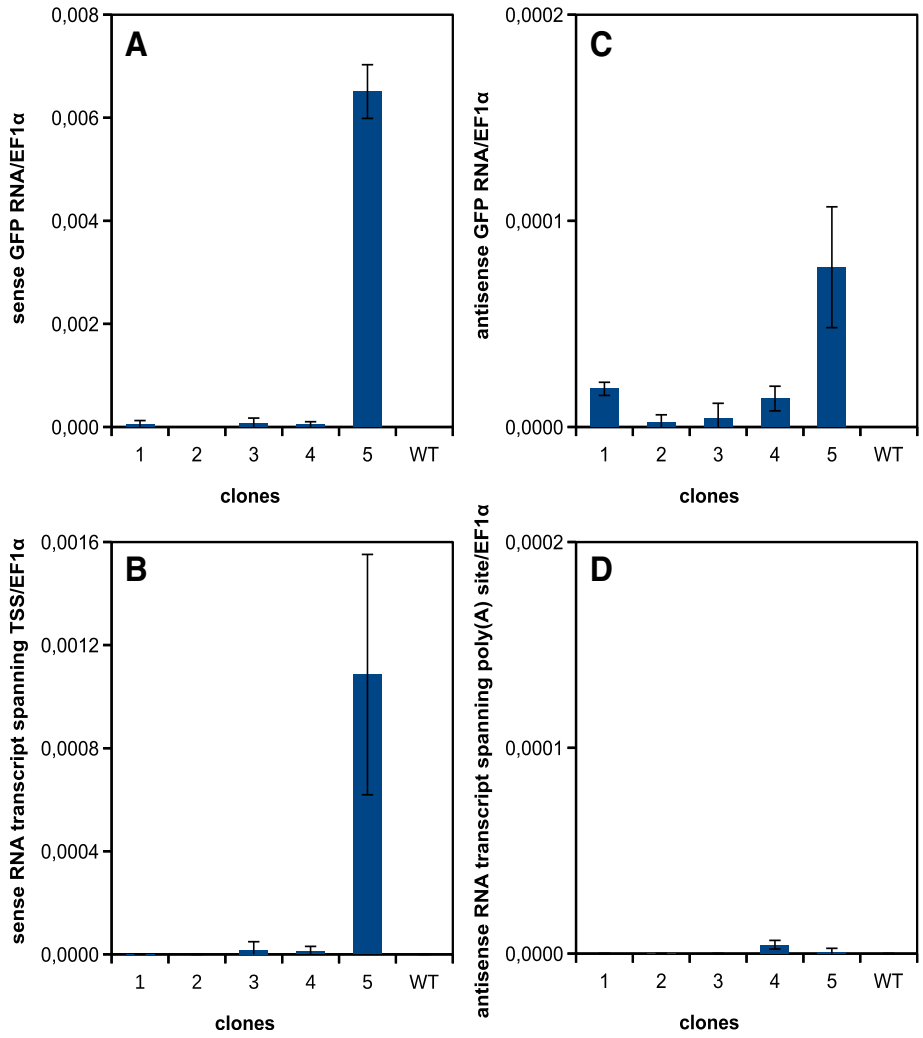


Fig. 7 Transcription of *GFP* gene in siliques of selected *Arabidopsis thaliana* transformants with *pER8-GFP*. RT-qPCR analysis of transcript levels in five selected plants untreated with β-estradiol. **a** The level of the sense *GFP* transcript (RT: R1 primer, qPCR: F1 and R1 primers); **b** the level of the sense transcript containing the region 50 nt upstream of transcription start site (TSS) of the inducible promoter (RT: R1 primer, qPCR: F2 and R2 primers); **c** the level of the antisense *GFP* transcript (RT: F1 primer, qPCR: F1 and R1 primers); **d** the level of the antisense transcript containing the region 50 nt downstream of the last poly(A) signal of T_{3A} terminator (RT: F1 primer, qPCR: F3 and R3 primers)

sequence itself was the cause of the spontaneous transcription. The *IR-PIN3* was also clearly transcribed in vast majority of analyzed calli, although we did not observe strong *PIN3* silencing as in the case of *GFP* (Fig. 4). Repeat regions are often targets of silencing [13, 41], so we asked whether the repetitive nature of our silencing inducers (*IR-GFP* and *IR-PIN3*) could be the reason for why they were spontaneously transcribed. But RT-qPCR analysis showed that single sense *GFP* gene was transcribed at similar frequency and level in BY-2 calli, so we assumed that neither the repeat structure of the sequence was necessary for the spontaneous transcription (Fig. 5). Since equal frequency of *GFP* silencing was observed with two different *T-DNAs* carrying the *IR-GFP*, we could conclude that the spontaneous transcription was unlikely connected with any regulatory elements present within the *T-DNAs*.

Expression of endogenous *PIN3* was more resistant against silencing induced by spontaneous transcription of *IR-PIN3*

Unlike *IR-GFP*, spontaneous transcription of *IR-PIN3* resulted in only small decrease or unaffected level of *PIN3* mRNA (Fig. 4 and Additional file 2). The level of *IR-PIN3* transcription was somewhat lower than the transcription of *IR-GFP*, but the lowest transcript level of *IR-GFP* able to silence *GFP* expression was lower than the highest transcript level of *IR-PIN3* that was not able to silence *PIN3* expression (Figs. 3 and 4). The *IR-PIN3* transcript could be theoretically less efficient in forming dsRNA and producing siRNAs. However, after treatment with β-estradiol, the same construct worked as an effective inducer of silencing (Jan Petrášek, personal communication). An alternative explanation is that the difference was related to the higher sensitivity of artificially introduced *GFP* transgene

to silencing. This could be connected with the absence of introns in the *GFP* gene. The presence of an intron and its splicing can suppress production of secondary siRNAs that might be necessary to amplify silencing when the transcript level of the inducer is low [42–44].

Pol II was involved in the spontaneous transcription

Plant specific RNA Polymerases IV and V provide wide-spread transcription of genomic DNA important for de novo DNA methylation and maintenance of heterochromatin [45–47]. The transcripts we detected thus may originate from these polymerases. However, the size of the transcripts we detected exceeded the size typical for Pol IV and V transcripts, i.e. tens of nt long up to few hundreds nt respectively [48, 49]. Moreover, at least some transcripts in the sense direction were spliced (Fig. 3g), which is typical for Pol II transcription. The ratio between the spliced and unspliced transcripts was similar in calli with spontaneous transcription and in estradiol-induced control line, where transcription was fully done by Pol II (compare columns e in Fig. 3c and g). The very low levels of spliced transcripts likely resulted from instant processing by the DCLs [13]. Amplification of transcripts from cDNA samples prepared by reverse transcription with oligo dT primers further indicates that at least a portion of the transcripts were polyadenylated (Additional file 5). Although we cannot exclude the possibility that other polymerases contributed to the spontaneous transcription, Pol II surely participated, because splicing and the presence of a polyA tail are characteristic features for its transcription.

Pervasive character of spontaneous transcription was possibly connected with specific chromatin state of some *T-DNAs*

The level of the spontaneous transcription strongly differed between independently transformed lines carrying the same *T-DNA*, indicating that the spontaneous transcription was not general, but rather specific to certain insertion events/loci.

Our attempts to find the transcription start site indicated that it was located outside of the *T-DNA*. It was previously demonstrated that transcripts originating from adjacent genomic regions could read-through across the border regions of *T-DNA* inserts and affect expression of transgenes present near the border sequence [50, 51]. Such transcription can also originate from a neighboring *T-DNA*. For this to occur in our system, the *T-DNAs* would have to be arranged either as tandem direct repeats or as tail-to-tail inverted repeats. But, it seems unlikely, that two thirds of transformed lines would contain such *T-DNA* arrangement and also our Southern hybridization did not indicate increased presence of such arrangements among the silenced lines. Therefore, we hypothesize that it was just the higher *T-DNA* copy number detected in silenced lines, which increased the probability that at least one *T-DNA* copy was inserted in a genomic region supporting the spontaneous transcription.

In the case of sense transcripts spanning the *pER8 T-DNA*, the polymerase was able to run across two terminators, yet the frequency of spontaneous transcription (silencing) was the same as in the case of the *pGreen T-DNA*, where *IR-GFP* was located near the border sequence. It should be also noted that the spontaneous transcriptional activity was often bidirectional, and that there was a correlation between the sense and antisense transcription levels in many calli. Based on these observations, we hypothesize that the spontaneous transcription was connected with specific chromatin state in some *T-DNA* insertion loci.

Our previous study indicated that the establishment of epigenetic marks can be accidental and at least in some cases independent of the chromosomal environment, since different epigenetic states could be established in the same insertion locus [24]. *T-DNAs,* when being inserted into the chromosome, are likely free of any epigenetic marks; as such, the formation of new marks may either reflect the chromatin state in the insertion locus or the new chromatin may be labelled with marks specific for new DNA insertions. This could make it prone to basal transcription by polymerase II that can somehow ignore terminators, thus forming long transcripts, irrespective of canonical transcription units present in the sequence. It is known, that facultative termination of long non-coding RNAs (lnRNAs) can be involved in regulation of gene expression and silencing [52–54] and that the chromatin state does have an effect on transcription termination [55]. More specifically, it was demonstrated that IBM2 allows Pol II to read-through silenced transposable elements inserted in introns of genes [56] and recent report showed that similar mechanism can also work for *T-DNAs* in introns of genes [57]. So we hypothesize that this long read-through transcription of newly integrated *T-DNAs* could be a component of genomic safety mechanisms evolved to allow silencing of newly integrated invasive DNAs (like transposable elements, TE). Such long polymerase II transcripts could be recognized by ubiquitous TEs-derived hc-siRNAs to initiate their silencing. Being an internal part of long transcript, there would be a little risk of translation of the TE transcript into executive proteins that would activate their replication/transposition.

Read-through transcription of inserted *T-DNAs* might be a more general phenomenon

The important question is how widespread and how significant this phenomenon can be. We tested three types of gene constructs (*IR-GFP*, *IR-PIN3* and *GFP*) within

two different *T-DNAs* (*pER8* and *pGreen*) using two different model organisms (BY-2 tobacco cell line and *Arabidopsis thaliana* plants). The results with various *T-DNAs* in BY-2 tobacco cell line were highly consistent, showing high frequency of read-through transcription and relatively high transcript levels. Additional analysis done with one *T-DNA* in *Arabidopsis thaliana* slightly differed, showing lower levels of transcription and lower occurrence among analyzed transformed plants. This could be connected with the physiological state of cells used for analyses (cell line vs. intact plant tissues) or different transcriptional regulation in species with dissimilar genome size (tobacco vs. *Arabidopsis thaliana*).

We also searched previous studies, which used inducible systems in plants to find further support for our observation. But most of the studies did not report the levels of RNA and if yes, then it is unclear if the transcript was not detected due to its absence or because it was below the detection limit. For example, Dohi et al. [23] did not detect any transcripts from the uninduced XVE system in BY-2 cell lines using Northern blot with DIG-labeled probes, but neither we were able to detect this transcription using the same detection method (data not shown), yet the RT-qPCR method clearly confirmed presence of these transcripts. In accord with our results, Kubo et al. [58] also detected low levels of transcripts from uninduced XVE system in *Physcomitrella patens* using RT-qPCR.

Additional strong support comes from early studies on silencing that used promoterless constructs as controls; they analyzed if silencing was caused by RNA or if it could be mediated directly by DNA-DNA interaction [11, 51, 59–62]; these works were done in *Petunia*, tobacco, *Arabidopsis thaliana* and *Neurospora*. Of the works mentioned above, only Cogoni et al. [51] partially characterized the origin of the transcripts causing the silencing in *Neurospora crassa*, by showing that they likely originated outside of the transgene. Whereas some studies did not observe silencing with promoterless constructs [63, 64], several other studies indeed reported induction of silencing by constructs or *T-DNAs* arranged as inverted repeats even without the presence of a promoter sequence [11, 59, 60]. All these old observations could be easily explained as read-through pervasive transcription of inverted repeats that produced hairpin RNAs.

Our data illustrate high incidence of pervasive read-through transcription of *T-DNAs* in BY-2 cells. These data are supported by very large numbers of analyzed independently transformed lines, 200 to 800 calli per each variant. In plants, only Sijen et al. [60] previously showed transcripts that likely originated from the silencer locus. As such, our report provides a considerable contribution to a more than fifteen years unsolved enigma, which has been experimentally overlooked with exception of some studies on DNAi [65]. We presume that the phenomenon of read-through transcription of *T-DNAs* is general, but specifically manifests only when working with *IR*, which has high potential to effectively induce silencing [14]. In contrast, low-level transcription of other constructs easily passes unnoticed, as the long-range character of transcripts disables translation into functional proteins that could visibly affect the phenotype.

Conclusions

We observed unexpectedly high frequency of read-through low-level transcription of several different *T-DNAs* in tobacco BY-2 cell lines and to some extent also in *Arabidopsis thaliana*. We show that this transcription was at least partially catalyzed by Pol II, which was able to read-through two standard terminators. We speculate that this unusual transcription was connected with establishment of specific chromatin state in some *T-DNA* insertions. Such read-through transcription could be for example a component of a safety mechanism for recognition and silencing of invasive DNA insertions. In the case of *T-DNAs* containing *IR-GFP*, the spontaneous read-through transcription was sufficient to initiate very efficient silencing of the *GFP* gene (*in trans*) in the majority of analyzed BY-2 calli. From the practical view, it is important that the researchers using inducible silencing systems should be aware of this phenomenon as it can in some cases largely affect the obtained results.

Additional files

Additional file 1: List of primers used in this study. (XLS 13 kb)

Additional file 2: Comparison of mean transcript levels from the analyzed T-DNAs. (A) Means of the IR transcript levels presented in Fig. 3 and Fig. 4. The means were calculated separately for the spontaneously silenced independent calli (i.e. 5 biological replicates, marked -) and non-silenced calli (marked +), one additional non-silenced callus grown on the induction medium with β-estradiol (marked as "e") and on the control medium with DMSO (marked as "d") is presented alongside. For the IR-PIN3, the category "spontaneously silenced calli" include half of the calli with the lowest expression of *PIN3* mRNA, the rest of the calli is part of the "non-silenced calli" category. Note that the higher expression of *IR-PIN3* in the category "spontaneously silenced calli" is not statistically significant. (B) Means of the transcript levels presented in Fig. 5. In all the experiments, intron from the inverted repeat (for both *IR-GFP* and *IR-PIN3*) is amplified with the same set of primers, so direct comparison of the transcript levels is possible. Also all the qPCR data are corrected for PCR efficiency (see Methods), so approximate comparison of quantities for different transcripts is also possible. The error bars represent standard deviations. (PDF 63 kb)

Additional file 3: Southern hybridization of total genomic DNA from BY-2 calli transformed with *pER8-IR-GFP*. Silenced independent calli marked with "-" and non-silenced calli marked with "+" (see Fig. 3). A DIG-labelled probe of the *HPT* gene was hybridized with DNA cleaved by *Nsi*I (N) and *Ase*I (A). *T-DNA* copy number was estimated as the number of hybridizing bands. The presence of repeats was analyzed as follows: tandem direct repeat should give 6.8 kbp fragment with both *Nsi*I and *Ase*I, plus one fragment of unknown size for *Nsi*I and *Ase*I; head-to-head inverted repeat should give 7.3 kbp fragment when digested with *Nsi*I and two fragments of unknown size when digested with *Ase*I; tail-to-tail inverted repeat should give 9.6 kbp fragment when digested with *Ase*I

and two fragments of unknown sizes when digested with *Nsi*I. (PDF 341 kb)

Additional file 4: Transcription of *GFP* gene in leaves of selected *Arabidopsis thaliana* transformants with *pER8-GFP*. RT-qPCR analysis of transcript levels in five selected lines untreated with β-estradiol. (A) The level of the sense *GFP* transcript; (B) the level of the sense transcript containing the region 50 nt upstream of transcription start site (*TSS*) of the inducible promoter; (C) the level of the antisense *GFP* transcript; (D) the level of the antisense transcript containing the region 50 nt downstream of the last poly(A) signal of T_{3A} terminator. (PDF 39 kb)

Additional file 5: Detection on polyA-tailed *GFP* transcripts in selected BY-2 calli transformed with *pER8-GFP*. Semiquantitative RT-PCR analysis of transcript levels in five independent calli untreated with β-estradiol (the same calli as in Fig. 5). cDNA was prepared using oligo *dT* primers. (A) The level of the *Actin* transcript (internal standard); (B) the level of the *GFP* transcript; (C) amplification of RNA samples that were not treated with reverse transcriptase to ensure that there was no DNA contamination. (PDF 41 kb)

Abbreviations
AS: antisense; IR: inverted repeat; LB: left border sequence of T-DNA; RB: right border sequence of T-DNA; UT: unterminated (without terminator)

Acknowledgments
We thank to Jan Petrášek for providing the *IR-PIN3* construct, to Miloš Duchoslav for the intron from *StPsbO* gene and to Adéla Růžičková for *Arabidopsis* plants transformed with an inducible *GFP* construct. We also thank to Kelly Hennessey for language corrections.

Funding
The work was supported by Ministry of Education, Youth and Sports of Czech Republic (project No. LO1417).

Availability of data and materials
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
VC carried out all the experimental work. VC together with LF designed the study, interpreted the results and wrote the manuscript. Both authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 October 2018 Accepted: 12 October 2018
Published online: 22 October 2018

References
1. Gelvin SB. Integration of agrobacterium T-DNA into the plant genome. Annu Rev Genet. 2017;51:195–217.
2. Kim S-I, Veena, Gelvin SB. Genome-wide analysis of agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. Plant J Cell Mol Biol. 2007;51:779–91.
3. Shilo S, Tripathi P, Melamed-Bessudo C, Tzfadia O, Muth TR, Levy AA. T-DNA-genome junctions form early after infection and are influenced by the chromatin state of the host genome. PLoS Genet. 2017;13:e1006875.
4. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science. 2003;301:653–7.
5. Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Körber H, Redei GP, et al. High-frequency T-DNA-mediated gene tagging in plants. Proc Natl Acad Sci. 1989;86:8467–71.
6. Hewelt A, Prinsen E, Schell J, Van Onckelen H, Schmülling T. Promoter tagging with a promoterless ipt gene leads to cytokinin-induced phenotypic variability in transgenic tobacco plants: implications of gene dosage effects. Plant J. 1994;6:879–91.
7. Hsing Y-I, Chern C-G, Fan M-J, Lu P-C, Chen K-T, Lo S-F, et al. A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol Biol. 2007;63:351–64.
8. Oltmanns H, Frame B, Lee L-Y, Johnson S, Li B, Wang K, et al. Generation of backbone-free, low transgene copy plants by launching T-DNA from the agrobacterium chromosome. Plant Physiol. 2010;152:1158–66.
9. Cluster PD, O'Dell M, Metzlaff M, Flavell RB. Details of T-DNA structural organization from a transgenic Petunia population exhibiting co-suppression. Plant Mol Biol. 1996;32:1197–203.
10. Kregten M van, Pater S de, Romeijn R, Schendel R van, Hooykaas PJJ, Tijsterman M. T-DNA integration in plants results from polymerase-θ-mediated DNA repair. Nat Plants 2016;2:16164.
11. Stam M, De Bruin R, Kenter S, Van Der Hoorn RAL, Van Blokland R, Mol JNM, et al. Post-transcriptional silencing of chalcone synthase in Petunia by inverted transgene repeats. Plant J. 1997;12:63–82.
12. Fojtová M, Bleys A, Bedřichová J, Houdt HV, Křížová K, Depicker A, et al. The trans-silencing capacity of invertedly repeated transgenes depends on their epigenetic state in tobacco. Nucleic Acids Res. 2006;34:2280–93.
13. Bologna NG, Voinnet O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. Annu Rev Plant Biol. 2014; 65:473–503.
14. Yan H, Chretien R, Ye J, Rommens CM. New construct approaches for efficient gene silencing in plants. Plant Physiol. 2006;141:1508–18.
15. Nagata T, Nemoto Y, Hasezawa S. Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. Int Rev Cytol. 1992;132:1–30.
16. Srba M, Černíková A, Opatrný Z, Fischer L. Practical guidelines for the characterization of tobacco BY-2 cell lines. Biol Plant. 2016;60:13–24.
17. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell. 2012;151:194–205.
18. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science. 2012;337:1360–4.
19. Lunerová-Bedřichová J, Bleys A, Fojtová M, Khaitová L, Depicker A, Kovařík A. Trans-generation inheritance of methylation patterns in a tobacco transgene following a post-transcriptional silencing event. Plant J. 2008;54: 1049–62.
20. Zhao M, San León D, Delgadillo MO, García JA, Simón-Mateo C. Virus-induced gene silencing in transgenic plants: transgene silencing and reactivation associate with two patterns of transgene body methylation. Plant J. 2014;79:440–52.
21. Butaye KMJ, Cammue BPA, Delauré SL, Bolle MFCD. Approaches to minimize variation of transgene expression in plants. Mol Breed. 2005; 16:79–91.
22. Dalakouras A, Tzanopoulou M, Tsagris M, Wassenegger M, Kalantidis K. Hairpin transcription does not necessarily lead to efficient triggering of the RNAi pathway. Transgenic Res. 2010;20:293–304.
23. Dohi K, Nishikiori M, Tamai A, Ishikawa M, Meshi T, Mori M. Inducible virus-mediated expression of a foreign protein in suspension-cultured plant cells. Arch Virol. 2006;151:1075–84.
24. Nocarova E, Fischer L. Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. BMC Plant Biol. 2009;9:44.
25. Moore I, Samalova M, Kurup S. Transactivated and chemically inducible gene expression in plants. Plant J. 2006;45:651–83.
26. Zuo J, Niu Q, Chua N. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. 2000;24:265–73.
27. Davis SJ, Vierstra RD. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol. 1998;36: 521–8.

28. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for agrobacterium-mediated plant transformation. Plant Mol Biol. 2000;42:819–32.
29. Smith NA, Singh SP, Wang M-B, Stoutjesdijk PA, Green AG, Waterhouse PM. Gene expression: Total silencing by intron-spliced hairpin RNAs. Nature. 2000; 407:319–20.
30. Murashige T, Skoog FA. Revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant. 1962;15:473–97.
31. Deblaere R, Bytebier B, De Greve H, Deboeck F, Schell J, Van Montagu M, et al. Efficient octopine Ti plasmid-derived vectors for agrobacterium-mediated gene transfer to plants. Nucleic Acids Res. 1985;13:4777–88.
32. Aranda PS, LaJoie DM, Jorcyk CL. Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis. 2012;33:366–9.
33. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55:611–22.
34. Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 2003;339:62–6.
35. Dvořáková L, Srba M, Opatrny Z, Fischer L. Hybrid proline-rich proteins: novel players in plant cell elongation? Ann Bot. 2012;109:453–62.
36. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (Nicotiana tabacum) during development and abiotic stress. Mol Genet Genomics. 2010;283:233–41.
37. Nocarova E, Opatrny Z, Fischer L. Successive silencing of tandem reporter genes in potato (Solanum tuberosum) over 5 years of vegetative propagation. Ann Bot. 2010;106:565–72.
38. Ji G, Li L, Li QQ, Wu X, Fu J, Chen G, et al. PASPA: a web server for mRNA poly(A) site predictions in plants and algae. Bioinforma Oxf Engl. 2015;31:1671–3.
39. Degenhardt RF, Bonham-Smith PC. Arabidopsis ribosomal proteins RPL23aA and RPL23aB are differentially targeted to the nucleolus and are disparately required for Normal development. Plant Physiol. 2008;147:128–42.
40. Kang C-Y, Lian H-L, Wang F-F, Huang J-R, Yang H-Q. Cryptochromes, Phytochromes, and COP1 regulate light-controlled stomatal development in Arabidopsis. Plant Cell. 2009;21:2624–41.
41. Sasaki T, Lee T, Liao W-W, Naumann U, Liao J-L, Eun C, et al. Distinct and concurrent pathways of pol II and pol IV-dependent siRNA biogenesis at a repetitive trans-silencer locus in Arabidopsis thaliana. Plant J. 2014;79:127–38.
42. Christie M, Croft LJ, Carroll BJ. Intron splicing suppresses RNA silencing in Arabidopsis. Plant J. 2011;68:159–67.
43. Vaistij FE, Jones L, Baulcombe DC. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. Plant Cell. 2002;14:857–67.
44. Vermeersch L, De Winne N, Depicker A. Introns reduce transitivity proportionally to their length, suggesting that silencing spreads along the pre-mRNA. Plant J. 2010;64:392–401.
45. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014;507:124–8.
46. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498:385–9.
47. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, et al. Spatial and functional relationships among pol V-associated loci, pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev. 2012;26:1825–36.
48. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, et al. Identification of pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. elife. 2015;4:e09591.
49. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase pol IVb/pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell. 2008;135:635–48.
50. Breyne P, Gheysen G, Jacobs A, Montagu MV, Depicker A. Effect of T-DNA configuration on transgene expression. Mol Gen Genet. 1992;235:389–96.
51. Cogoni C, Irelan JT, Schumacher M, Schmidhauser TJ, Selker EU, Macino G. Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J. 1996;15:3153–63.
52. Mischo HE, Proudfoot NJ. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim Biophys Acta BBA - Gene Regul Mech. 2013;1829:174–85.
53. Touat-Todeschini L, Shichino Y, Dangin M, Thierry-Mieg N, Gilquin B, Hiriart E, et al. Selective termination of lncRNA transcription promotes heterochromatin silencing and cell differentiation. EMBO J. 2017;36:2626–41.
54. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science. 2010;327:94–7.
55. Morse NJ, Gopal MR, Wagner JM, Alper HS. Yeast terminator function can be modulated and designed on the basis of predictions of nucleosome occupancy. ACS Synth Biol. 2017;6:2086–95.
56. Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, Ito T, et al. Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin. Nat Commun. 2013;4. https://doi.org/10.1038/ncomms3301.
57. Osabe K, Harukawa Y, Miura S, Saze H. Epigenetic regulation of Intronic transgenes in Arabidopsis. Sci Rep. 2017;7. https://doi.org/10.1038/srep45166.
58. Kubo M, Imai A, Nishiyama T, Ishikawa M, Sato Y, Kurata T, et al. System for stable β-estradiol-inducible gene expression in the Moss Physcomitrella patens. PLoS One. 2013;8:e77356.
59. Luff B, Pawlowski L, Bender J. An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. Mol Cell. 1999;3:505–11.
60. Sijen T, Vijn I, Rebocho A, van Blokland R, Roelofs D, Mol JNM, et al. Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr Biol. 2001;11:436–40.
61. Van Blokland R, Van der Geest N, Mol JNM, Kooter JM. Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. Plant J. 1994;6:861–77.
62. Voinnet O, Vain P, Angell S, Baulcombe DC. Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic Promoterless DNA. Cell. 1998;95:177–87.
63. Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJM. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. 2000;19:5194–201.
64. Waterhouse PM, Graham MW, Wang M-B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci. 1998;95:13959–64.
65. Kawai-Toyooka H, Kuramoto C, Orui K, Motoyama K, Kikuchi K, Kanegae T, et al. DNA interference: a simple and efficient gene-silencing system for high-throughput functional analysis in the Fern Adiantum. Plant Cell Physiol. 2004;45:1648–57.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

Chapter 25

Plant Cell Lines in Cell Morphogenesis Research: From Phenotyping to -Omics

Petr Klíma, Vojtěch Čermák, Miroslav Srba, Karel Müller, Jan Petrášek, Josef Šonka, Lukáš Fischer, and Zdeněk Opatrný

Abstract

Here we provide an overview of procedures for long-term cultivation, phenotyping, genotyping, and genetic transformation of cell cultures of tobacco cell lines BY-2 and VBI-0, and of *A. thaliana*, ecotype Landsberg erecta (LE) cell line. Notably, we present an improved protocol for BY-2 transformation and cloning and extend the available plant cell lines methodology toward high-throughput technologies like fluorescent-based cell sorting and transcriptomics.

Key words Tobacco, Cell lines, BY-2, VBI-0, Transformation, Cloning, Cell sorting, Transcriptomics

1 Introduction

Plant cell lines represent a unique model for the study of the basic processes of plant morphogenesis, that is, cell division, growth and development, including intracellular functional and structural differentiation, as well as shape changes of cells and their aggregates. They allow for characterizing their changing phenotype simultaneously with a detailed biochemical and molecular analysis. The results obtained can then be interpreted in the context of morphogenesis of the whole organism (for details *see* refs. 1, 2).

Our previous publication [3] presents a series of basic protocols for working with two model tobacco cell lines, VBI-0 (established in 1967; cf. refs. 4–6) and BY-2 [7], and one Arabidopsis model line, ecotype Landsberg erecta (LE). Some procedures are then elaborated on in our recently published methodological paper [8]. Both these publications deal mainly with basic cell phenotyping (preparation and cultivation of lines, determination of their growth curves, detection of viability, and description of changes in the micromorphology of cultures, especially in the context of cell growth polarity and division).

Fatima Cvrčková and Viktor Žárský (eds.), *Plant Cell Morphogenesis: Methods and Protocols*, Methods in Molecular Biology, vol. 1992, https://doi.org/10.1007/978-1-4939-9469-4_25, © Springer Science+Business Media, LLC, part of Springer Nature 2019

The paper by Seifertová et al. [3] also contains a transformation protocol, the revised version of which makes part of this chapter. An efficient and rapid preparation of transgenic cell lines allows to study the function of inserted foreign genes and to use these models for functional genomics. Such a strategy serves as an alternative to the very limited possibilities of establishing fast-growing, friable and phenotypically stable cell cultures of other plant species. While the biological causes for the many years of failure in their derivation remain unclear (*see* ref. 2), our revised transformation protocol for BY-2 may help to mitigate at least some of these shortcomings.

Cell colonies obtained after transformation can be genetically or epigenetically heterogeneous (especially when using CRISPR/CAS9 genome editing tool), which seriously complicates subsequent analyses. In addition, frequently, the expression of introduced genes is spontaneously silenced in certain subpopulations of the transgenic lines. Our simple method of cloning of individual transformed cells allows for overcoming this complication [9].

Cell cultures also allow precise and simple quantification of fluorescent signals in huge populations (tens of thousands) of individual cells. The fluorescent signals can originate from fluorescent dyes (e.g., allowing to determine DNA content) or reflect the level of GFP or other fluorescent proteins (e.g., allowing to study gene silencing or sensing of auxin level with fluorescent proteins controlled by the DR5 promoter).

Thanks to the above procedures performed under precisely defined in vitro conditions, we can study the mechanisms of cellular morphogenesis regulated by both the "intracellular" signals and signals providing immediate or distant intercellular communication. The model system of cell lines characterized in detail (de facto cell strains—*see* ref. 1) further represents a sensitive tool to investigate the effect of various stress factors, xenobiotics, or pathogens on a plant cell or organism.

Information on the genetic background of the studied models usually significantly facilitates the mapping of the observed phenotypic changes to their molecular biological backing. Although numerous profile conferences have already been organized in the name of the BY-2 cell line and two monographs have been published [10, 11], knowledge of the genome, transcriptome, and proteome of BY-2 cell line remains largely incomplete to date. As a contribution in this regard, we provide simple directions on how to obtain transcriptomic data of BY-2 and VBI-0 cell lines using a servicing company; then we suggest a pipeline to assess the data and present a "reference" transcriptomic pattern of both cell cultures. Altogether we believe that the advances in the omics field together with the possibility of targeted modifications on purpose [12, 13] can sustain the future relevance of the plant cell lines as unique experimental models in plant biology and biotechnology-oriented research.

2 Materials

2.1 For BY-2 Transformation

1. BY-2 cell line (*see* **Note 1**), *Agrobacterium tumefaciens* (strains, e.g., C58C1, GV2260, LBA115, or others) carrying a binary vector of interest.
2. YEB medium for *Agrobacterium tumefaciens* cultivation [14].
3. 20 mM acetosyringone (1000× stock in ethanol).
4. Three 100 ml Erlenmeyer flasks with 30 ml sterile liquid MS medium modified for BY-2 cell culture according to ref. 7.
5. Selection plates: 6 cm plastic petri dishes with agar solidified MS medium (*see* above), supplemented with selection antibiotics or herbicide and 100 mg/l cefotaxime (from 1000× aqueous stock). Each petri dish is filled with 11–13 ml of selection medium; about ten petri dishes is recommended for every transformation variant.
6. Sterile cell filtration set (e.g., Nalgene "Filter Holder with Receiver", or any similar instrument) equipped with 20 μm nylon mesh filter for washing out *Agrobacterium* cells after cocultivation.
7. Common equipment for sterile tissue culture handling.

2.2 For Cloning of Transgenic BY-2 Lines

1. Wild type (WT) and antibiotic/herbicide-resistant transgenic BY-2 cell lines.
2. Horizontal shaker, laminar hood.
3. Standard liquid and solidified cultivation media w/o antibiotic (herbicide).

2.3 For Flow Cytometry

1. BY-2 cell line (expressing fluorescent protein gene or labeled with a fluorescent dye).
2. Horizontal shaker.
3. Centrifuge.
4. BD LSR II flow cytometer or other appropriate flow cytometer.
5. Enzyme solution: 0.5% (w/v) cellulase R-10 (Duchefa), 0.1% (w/v) pectolyase Y-23 (Duchefa), and 0.45 M mannitol (Sigma-Aldrich; *see* **Note 2**).
6. MS medium with 0.4 M sucrose. This medium can be prepared by adding 1.07 g of sucrose to 10 ml of the standard MS medium for BY-2 cultivation.

2.4 For Transcriptomics

1. BY-2 cell line from a particular phase of the subculture interval or carrying a particular gene of interest or subjected to any other condition of interest.
2. Horizontal shaker, laminar hood.

3. Sterile cell filtration set (e.g., Nalgene "Filter Holder with Receiver", or any similar instrument) equipped with 20 μm nylon mesh filter.
4. Common equipment for sterile tissue culture handling.
5. Mortar and pestle, liquid nitrogen.
6. RNA isolation kit.
7. DNA removal kit.

3 Methods

3.1 BY-2 Transformation Protocol

According to our observation, the 3 days usually recommended for cocultivation with *Agrobacterium* are very stressful for the tobacco cells. Therefore, we shortened the cocultivation to just a single day and supplemented the cocultivated cell mixture with fresh untransformed (i.e., unstressed) BY-2 culture. This facilitates the transformed cells to start rapid growth and contributes to the rescue of transformed calli if the efficiency is very low, for example, due to a problematic binary vector or a harmful transgene construct (*see* **Note 3**).

1. Preparation of the cultures: (a) Inoculate a fresh single colony of *Agrobacterium tumefaciens* into 20 ml of YEB medium supplemented with appropriate selection antibiotics and cultivate overnight with shaking at 27–28 °C to late exponential (milky) stage; (b) Inoculate two 100 ml Erlenmeyer flasks containing 30 ml of sterile MS medium with 10 ml inoculum of 4- to 5-day-old BY-2 culture or 6 ml inoculum of 6- to 10-day-old culture. Cultivate under standard cultivation conditions overnight.
2. Cocultivation: (a) Add 40 μl of acetosyringone to one flask of BY-2 culture and pipet the culture up and down 20 times through 10 ml uncut pipette tip (diameter of the tip hole about 1 mm) in order to cause mechanical wounding to the tobacco cells (*see* **Note 4)**. (b) Add 600 μl of *A. tumefaciens* suspension to 6 ml of BY-2 culture. Pipet 2 ml aliquots of the mixed culture into separate sterile 6 cm petri dishes and cocultivate for another 24–30 h in darkness at 27 °C without shaking.
3. Washing the culture: (a) Dilute the second unused flask of BY-2 culture (from **step 1**) by adding 30 ml of fresh sterile MS medium and 65 μl of cefotaxime; (b) Collect the cocultivation mix by 10 ml pipette from petri dishes and wash 2–3 times by diluting in 50 ml of 3% sucrose and subsequent filtering in order to reduce *A. tumefaciens* density; (c) Resuspend the

washed filtered cells in 12 ml of the diluted untransformed BY-2 culture prepared in (a).

4. Selection: Transfer the washed and diluted culture into 6 cm petri dishes with solid selection media in two different volume aliquots: 0.75 ml and 1.5 ml (e.g., 12 ml can be distributed as 6 × 0.75 ml and 5 × 1.5 ml; **Note 5**).

3.2 Cloning of Transgenic BY-2 Lines

1. Dilute 2 ml of each transgenic stationary suspension culture (*see* **Note 6**) with 6 ml of fresh MS medium in a sterile tube.
2. Dilute 10 ml of WT stationary suspension culture with 30 ml of fresh MS medium in a sterile tube.
3. Pipet 4 μl of diluted transgenic culture with widely cut tip into 4 ml of WT stationary suspension culture in a new tube (1:1000 dilution). Mix by gently shaking.
4. Plate 0.5 ml of mixed cell culture onto a petri dish (∅ 6 cm) with solidified MS medium containing standard concentration of antibiotic (herbicide) used for selection.
5. Small calli representing clones of individual cells/cell files appear within 2 weeks (approximately 25 per plate).

3.3 Assessing Fluorescence of BY-2 Protoplasts Using Flow Cytometry

1. Take up to 100 mg of cells from suspension culture or callus (*see* **Note 7**) and place them into the 6-well cell culture plate (*see* **Note 8**). Add 1.5 ml of the enzyme solution, place the multiwell plate on an orbital shaker, and shake at 100 RPM for 2–3 h at 26 °C (or room temperature).
2. Transfer the enzyme solution with protoplasts (*see* **Note 9**) to a 2 ml microcentrifuge tube using a pipette with a cut tip. Centrifuge at 200 × *g* for 5 min (turn off the centrifuge brake if possible). Discard the enzyme solution (*see* **Note 2**) and gently resuspend the pellet in 600 μl of MS medium with 0.4 M sucrose.
3. Pipet 240 μl of the sample (*see* **Note 10**) to 96 multiwell plate using a pipette with a cut tip and analyze the samples on LSR II or other appropriate cytometer. The live protoplast population can be easily distinguished on the SSC-FSC dot plot (Fig. 1a, b, **Note 11**). At least 5000 live cells should be measured to get a smooth histogram of protoplast fluorescence (Fig. 1b).
4. Analyze the data using appropriate software, for example, freely available Flowing Software, which is reliable, user friendly and able to process larger number of samples easily (*see* **Note 12**). Output of such analyses can be for example average fluorescence of a cell or percentage of fluorescently positive cells (in such cases it is necessary to measure WT BY-2 cells to know the background fluorescence).

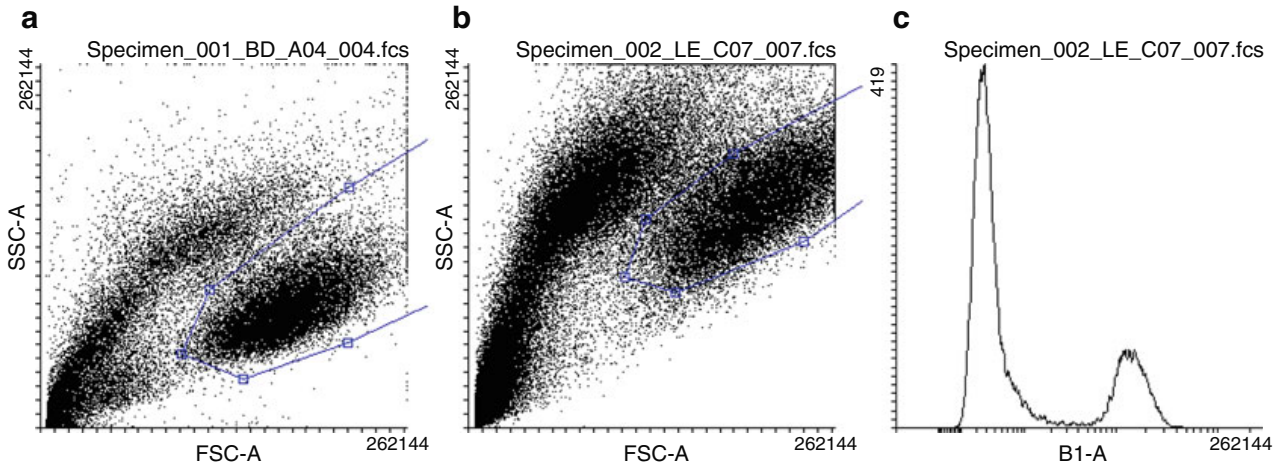


Fig. 1 Assessing fluorescence of BY-2 protoplasts using flow cytometry. (**a** and **b**) Examples of SSC-FSC dotplots, the live cell population can be easily distinguished—here it is in the blue gate; (**c**) an example of histogram of protoplast fluorescence in cell line with a subpopulation of cells that partially silenced expression of GFP transgene

3.4 Transcriptomics

1. RNA isolation**:** Collect up to 100 mg of suspension culture cells and freeze them immediately in liquid nitrogen (*see* **Note 13**). Homogenize the material in liquid nitrogen using mortar and pestle (*see* **Note 14**) and isolate the total RNA (*see* **Note 15**). Elute the RNA in RNase-, DNase-, and protease-free water or Tris buffer. Remove contaminating DNA (*see* **Note 16**).
2. Confirm RNA quality, purity, and concentration by 0.8% agarose gel electrophoresis, by measuring absorbance at 260 and 280 nm and by evaluation of the RNA Integrity (RIN) value on Agilent 2100 Bioanalyzer (*see* **Note 17**).
3. Send the samples to servicing company (GATC-biotech/Eurofins Genomics) where further sample processing (poly-A containing mRNA purification, fragmentation and preparation of strand-specific cDNA library) is performed. For simple quantification of transcripts, at least 15 million of 50 bps long single reads are desirable. Map the reads against *Nicotiana tabacum* v4.5 CDS database (ref. 15; **Note 18**). For assessment of general transcription profile similarity, principal component analysis (PCA) can be used (Fig. 2).

4 Notes

1. This rapid transformation protocol requires good fitness of the starting BY-2 cell culture. The culture needs to be able to multiply its biomass at least 20 times during week subculture interval (SBI), reaching 2.5×10^6 cells/ml at the end of SBI, and needs to reach at least 5% mitotic index on the second day after inoculation (*see* Subheading 3.1, **step 1**). If such growth

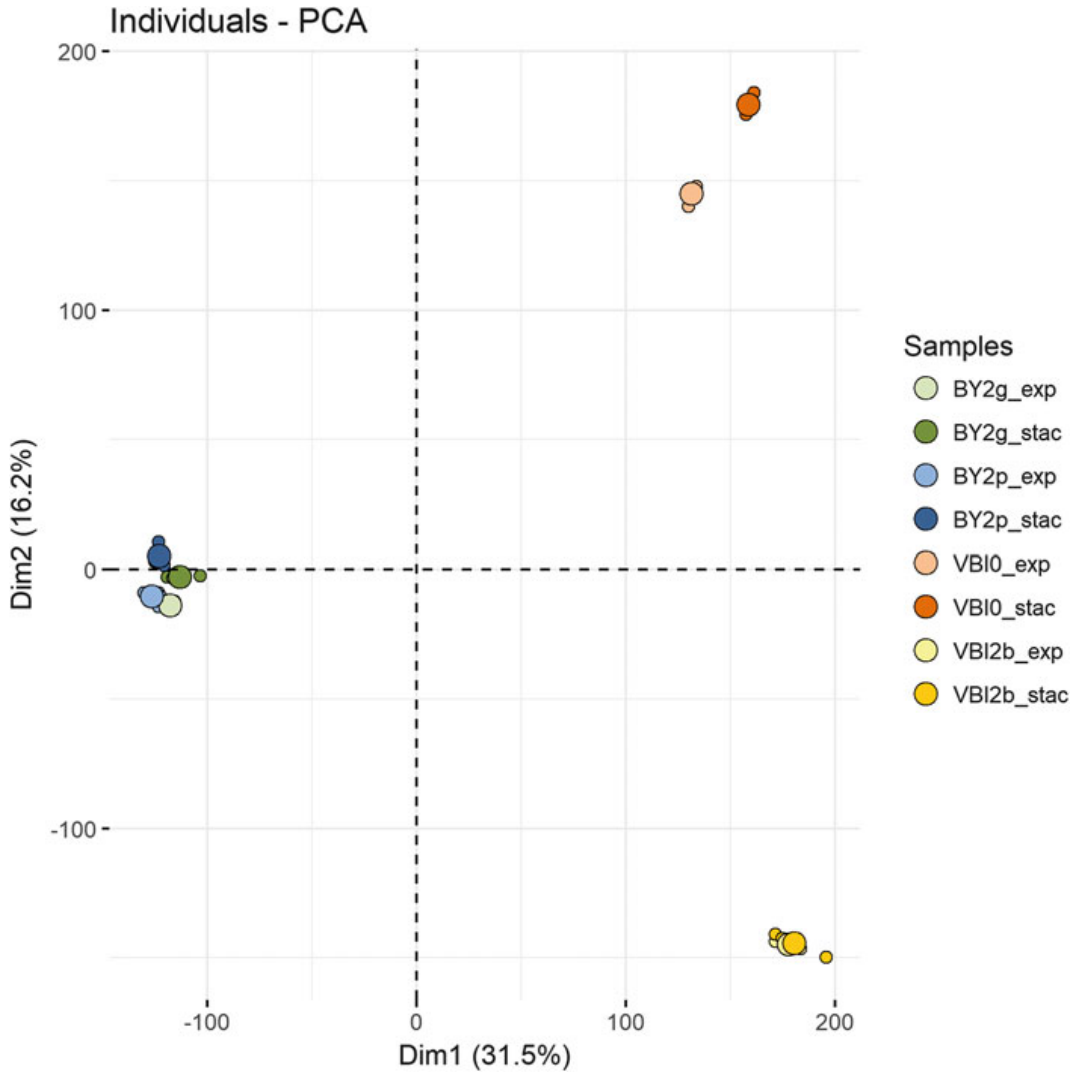


Fig. 2 Principal component analysis of RNAseq data for two lines of BY-2 and VBI tobacco cell cultures collected at exponential (2-days-old cells) and stationary (7-days-old cells) phases of the subculture interval. Each biological sample is represented by three independent replicates. Both exponential and stationary cells show characteristic sets of transcripts, providing evidence for the stability of the transcriptomes. *Exp* exponential, *stac* stationary, *BY2g, BY2p, VBI0, VBI2b* independently grown tobacco BY-2 and VBI cell lines

parameters are not reached, the cultivation protocol needs to be optimized rigorously according to ref. 7.

2. Prepared enzyme solution can be stored in freezer at −20 °C. After use, the solution can be recycled by filtering through 20 μm nylon mesh filter. It can be reused up to three times without obvious loss in protoplasting efficiency.
3. We also realized that it is not necessary to have exactly a 3-day--old exponential culture, just the sufficient mitotic index reaching at least 5% is necessary. Researchers are therefore free to start the protocol any day with late exponential or stationary culture, which can reach good mitotic index just 1 day after the initial subculture. The original transformation protocol is reduced from 7 days just to 3 days. Moreover, the growth of

transformed calli is typically faster, allowing the first subculture as soon as after 3–4 weeks, and the numeric yield of transformants is usually higher.

4. The culture can be divided into more transformation variants in this step; 6 ml of treated BY-2 culture is required for each transformation variant, so one flask (36 or 40 ml) can be used for transformation of up to six different constructs.
5. The transformation efficiency is commonly variable. When the two size aliquots are used, at least one of the variants gives appropriate density of transformed calli (10–50 calli per plate, which is neither too low nor too dense to complicate proper separation of individual calli). Typically, 6 ml of cocultivation mix provides 150 to 600 transformed calli on 11 selection plates. Treated volumes can be individually adjusted according to the number of required transformed calli, when keeping the ratios of transformed cells, untreated culture and fresh medium constant.
6. BY-2 lines in the callus form can be gently resuspended in liquid medium, cultivated on a rotor shaker for a week and subcultured once (approximately 1.5–3 ml of suspension into 30 ml of fresh medium) before cloning.
7. Starting from the suspension culture is better as the efficiency of protoplasting cells from calli is quite low. The medium from the suspension culture has to be removed. It can be done either by centrifugation or by filtration. For protoplasting we usually use suspension culture in the exponential phase of growth (3–4 days old).
8. Other types of containers can be used, if they allow adequate shaking.
9. The protoplasts can be checked by microscope; however, do not cover the sample directly with a cover glass, this would crush the protoplasts. After the incubation with enzymes, all cells should be spherical, some cells can by still stick together in files, but this is not a problem for the following flow cytometry.
10. In theory, dead protoplasts should sediment on the bottom of the microcentrifuge tube if they stay in 0.4 M sucrose. However, in our hands in some of the cell lines the live protoplasts sediment as well. Therefore, we recommend sampling the cells indiscriminately from the whole volume.
11. FDA staining as described in ref. 3 can verify the live cell population.
12. Flowing Software can be downloaded from http://flowingsoftware.btk.fi/.
13. Can be stored in −80 °C before further processing.

14. General rules for RNA work apply. It is advised to remove RNases from all accessories used. RNase ZAP (Sigma-Aldrich) is used.
15. RNeasy Plant Isolation kit (Qiagen) is used.
16. DNA-free kit (Ambion) is used.
17. Requirements for quality and quantity of RNA may depend on sequencing company. For example, GATC Biotech requires 1 μg of RNA of concentration >20 ng/μl, OD 260/280 between 1.8 and 2.2, RIN value >8.
18. Bioinformatics processing can be part of the sequencing contract. Several workflows exist; for more information see for example ref. 16.

Acknowledgments

The work was supported by the Czech Ministry of Education, Youth and Sports (project LO1417), Charles University Grant Agency (GAUK 1002216), and the Czech Science Foundation (projects GA16-10948S and GA16–19557S).

References

1. Opatrný Z (2014) From Němec and Haberlandt to plant molecular biology. In: Nick P, Opatrný Z (eds) Applied plant cell biology. Springer, Berlin, pp 1–36
2. Opatrný Z, Nick P, Petrášek J (2014) Plant cell strains in fundamental research and applications. In: Nick P, Opatrný Z (eds) Applied plant cell biology. Springer, Berlin, pp 455–481
3. Seifertová D, Klíma P, Pařezová M, Petrášek J, Zažímalová E et al (2014) Plant cell lines in cell morphogenesis research. Methods Mol Biol 1080:215–229
4. Opatrný Z, Opatrná J (1976) The specificity of the effect of 2,4-D and NAA on the growth, micromorphology, and occurrence of starch in long-term *Nicotiana tabacum* L. cell strains. Biol Plant 18:359–365
5. Březina V, Opatrný Z (1986) Analýza mikrokinematografického záznamu dělení a stárnutí buněk tabákového kmene. The analysis of microcinematographic record of the cell division and senescence of the tobacco cell strain (in Czech). Československá fyziologie 35:76
6. Opatrný Z (1986) Filamentous higher plant cell strains- a novel model for cell polarity studies. In: Abstr. 2nd European Cell Biology Congress, Budapest, p. 15
7. Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. Int Rev Cytol 132:1–30
8. Srba M, Černíková A, Opatrný Z, Fischer L (2016) Practical guidelines for the characterization of tobacco BY-2 cell lines. Biol Plant 60:13–24
9. Nocarová E, Fischer L (2009) Cloning of transgenic tobacco BY-2 cells; an efficient method to analyze and reduce high natural heterogeneity of transgene expression. BMC Plant Biol 9:44
10. Nagata T, Hasezawa S, Inzé D (eds) (2004) Tobacco BY-2 cells. Biotechnology in agriculture and forestry, vol 53. Springer-Verlag, Berlin
11. Nagata T, Matsuoka K, Inzé D (eds) (2006) Tobacco BY-2 cells: from cellular dynamics to omics. Biotechnology in agriculture and forestry, vol 58. Springer-Verlag, Berlin
12. Mercx S, Tollet J, Magy B, Navarre C, Bountry M (2016) Gene inactivation by CRISPR-Cas9 in *Nicotiana tabacum* BY-2 suspension cells. Front Plant Sci 7:40
13. Mercx S, Smargiasso N, Chaumont F, De Pauw E, Boutry M et al (2017) Inactivation of the β(1,2)-xylosyltransferase and the

α(1,3)-fucosyltransferase genes in *Nicotiana tabacum* BY-2 cells by a multiplex CRISPR/Cas9 strategy results in glycoproteins without plant-specific glycans. Front Plant Sci 8:403

14. Vervliet G, Holsters M, Teuchy H, Van Montagu M, Schell J (1975) Characterization of different plaque-forming and defective temperate phages in agrobacterium strains. J Gen Virol 26:33–48
15. Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD et al (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genomics 18:448
16. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13

Epigenetics & Chromatin

RESEARCH Open Access

Detailed insight into the dynamics of the initial phases of de novo RNA-directed DNA methylation in plant cells

Adéla Přibylová, Vojtěch Čermák, Dimitrij Tyč and Lukáš Fischer*

Abstract

Background: Methylation of cytosines is an evolutionarily conserved epigenetic mark that is essential for the control of chromatin activity in many taxa. It acts mainly repressively, causing transcriptional gene silencing. In plants, de novo DNA methylation is established mainly by RNA-directed DNA-methylation pathway. Even though the protein machinery involved is relatively well-described, the course of the initial phases remains covert.

Results: We show the first detailed description of de novo DNA-methylation dynamics. Since prevalent plant model systems do not provide the possibility to collect homogenously responding material in time series with short intervals, we developed a convenient system based on tobacco BY-2 cell lines with inducible production of siRNAs (from an RNA hairpin) guiding the methylation machinery to the *CaMV 35S* promoter controlling GFP reporter. These lines responded very synchronously, and a high level of promoter-specific siRNAs triggered rapid promoter methylation with the first increase observed already 12 h after the induction. The previous presence of CG methylation in the promoter did not affect the methylation dynamics. The individual cytosine contexts reacted differently. CHH methylation peaked at about 80% in 2 days and then declined, whereas CG and CHG methylation needed more time with CHG reaching practically 100% after 10 days. Spreading of methylation was only minimal outside the target region in accordance with the absence of transitive siRNAs. The low and stable proportion of 24-nt siRNAs suggested that Pol IV was not involved in the initial phases.

Conclusions: Our results show that de novo DNA methylation is a rapid process initiated practically immediately with the appearance of promoter-specific siRNAs and independently of the prior presence of methylcytosines at the target locus. The methylation was precisely targeted, and its dynamics varied depending on the cytosine sequence context. The progressively increasing methylation resulted in a smooth, gradual inhibition of the promoter activity, which was entirely suppressed in 2 days.

Keywords: Epigenetics, RdDM, RNA interference, sRNA sequencing, Transcriptional gene silencing

Background

All plant cells need to regulate gene expression in connection with developmental processes and as a reaction to external conditions. Simultaneously, the genetic information must be protected against invasive nucleic acids, mainly transposable elements (TEs). To avoid the detrimental effects of their activity, TEs must be kept inactive. However, TEs are integral components of genomes, frequently interspersed between functional genes, so cells need to differentially regulate the activity of particular regions within a genome [1]. For this purpose, cells possess a wide range of epigenetic tools for labelling chromatin at both the DNA and histone level. Histone labelling is highly complex, including a range of various posttranslational modifications of histone tails and varying representation of histone variants within nucleosomes, whereas DNA is labelled almost exclusively by methylation of cytosines (C) [2, 3].

*Correspondence: lukasf@natur.cuni.cz
Department of Experimental Plant Biology, Charles University, Faculty of Science, 128 44 Prague, Czech Republic

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Chromatin epigenetic marks are generally reversible, but they can also be very stable, especially in plants, where repressive marks are often transgenerationally inherited [4]. Therefore, their establishment has to be well-founded and highly specific. The sequence-specific chromatin repression can be realised either by the DNA-binding domains of transcription factors recruiting the polycomb repressive complex that induce trimethylation of H3K27 (these repressive marks are commonly reset between generations) [5] or by RNA-directed DNA methylation (RdDM). Target recognition in RdDM is based on the complementarity of small RNAs with nascent scaffold transcripts of plant-specific RNA polymerase V (Pol V) [3, 6].

DNA methylation serves as a repressive mark to inactivate gene transcription if it occurs in the promoter region [3, 7]; for TEs, methylation is usually spread along their full length [8]. In plants, DNA methylation is targeted on the C5 position of cytosines and can occur in any C contexts: CG, CHG, and CHH (where H can be A, C, or T). Once established, the methylation marks are maintained in dividing cells in three different ways. First, the methylation of CG is maintained by Methyltransferase 1 (MET1), which methylates C in hemi-methylated CG recognised by protein Variant in methylation 1 (VIM1). This process is tightly associated with DNA replication [9–11]. The other two mechanisms are based on the mutual connection between DNA methylation and histone posttranslational modifications—mainly H3K9me2. The first self-reinforcing loop is responsible for maintaining the methylation in CHG (and in heterochromatin also CHH) contexts by Chromomethylases (CMTs). These enzymes contain chromodomains that specifically bind to H3K9me2, which is likely needed for effective methylation of cytosines in the adjacent DNA [12, 13]. Vice versa, methylated CHG is recognised by SRA domain of histone methyltransferases Kryptonite (KYP, SUVH4) and SUVH5/6, which di-methylate H3K9 in the adjacent nucleosome(s) [14–20]. CHG context is maintained mainly by CMT3 [13, 21, 22]. CMT2 is responsible for methylation of CHH and also partly CHG context in canonical heterochromatin containing histone H1 [13, 22].

The last mechanism of "maintenance methylation" is most important for short TEs and border regions of long retrotransposons [8] and is based on the activity of Domains rearranged methyltransferase 2 (DRM2) in the process of canonical RdDM. RdDM is a part of the RNA interference (RNAi) machinery, which inactivates gene expression not only at the transcriptional (TGS) level, but also at the post-transcriptional (PTGS) level [23, 24]. In RdDM, DRM2 methylates C in a context-independent manner at *loci* that are complementary to small RNAs present in the cell [25]. There are several pathways of RdDM that have been described in plants in recent years. The canonical RdDM primarily serves to maintain CHH methylation in already repressed regions. It involves two plant-specific polymerases, Pol IV and Pol V [26–29]. Pol IV is responsible for the production of transcripts which serve as a source of small interfering RNAs (siRNAs) from genomic regions with inactive chromatin lacking histone H1, whereas Pol V assists in recognition of the target regions. Pol IV is attracted to chromatin via its interacting partner Sawadee homeodomain homologue 1 (SHH1), which binds to H3K9me2 and non-methylated H3K4 [30–32]. From *loci* with these chromatin labels, Pol IV creates 30–40-nt-long transcripts, which are replicated by RNA-dependent RNA polymerase 2 (RDR2) producing dsRNA precursors that are processed by Dicer-like 3 (DCL3) into 24-nt siRNA [33–36]. These siRNAs in association with Argonaute proteins AGO3/4/6 base-pair with nascent transcripts of Pol V and guide DRM2 for the methylation of C on the template DNA [6, 37, 38]. Pol V is primarily recruited to *loci* containing methylated cytosines via interaction with inactive SUVH homologs SUVH2/9 [39–41]. This canonical RdDM pathway serves not only to maintain methylation of genomic regions in *cis* but importantly, it should also allow siRNA-mediated "identity-based" recognition and de novo methylation/inactivation of newly inserted copies of TEs in *trans* [42, 43].

Recognition and de novo silencing of completely novel TEs (or transgenes) are likely expression-dependent and can be mediated by several other non-canonical RdDM pathways [43]. In addition to 24-nt siRNAs produced from transcripts of Pol IV in the canonical pathway, Pol II-dependent 24-nt and 21–22-nt siRNAs were also shown to be involved in the pathways, which are considered responsible for the methylation of DNA in *loci* not transcribed by Pol IV [3, 43]. In addition to typical siRNAs, recently discovered DCL-independent sidRNAs have also been suggested to be initial triggers of de novo DNA methylation of epigenetically naive *loci* [44], though a later study challenged this hypothesis [43].

While the molecular mechanisms of RdDM are relatively well-described at present, less is known about its dynamics. Voucheret already in 1994 showed that transcriptional *trans* silencing could start quickly in developing seed, but complete inactivation might require few weeks [45]. After massive leaf infiltration with Agrobacterium, rapid methylation of T-DNA was detectable in the promoter region just 2–3 day post-infiltration and the levels continued to rapidly accumulate over the 1st week and then steadily up to 21 days [46]. In mitotically dividing cells, the maintenance methylation of

newly synthesised DNA strands must be quick enough to ensure the replication of the epigenetic information between the two subsequent S-phases of the cell cycle. It is supposed to be exceptionally fast in the case of CG and CHG sequences. For instance, in human embryonic stem cells, the vast majority of the maintenance CG methylation takes place less than 20 min after replication [47]. However, to our knowledge, there is no information available for the dynamics of de novo RdDM initiation phases. Both a quick and slow model could apply; quick TGS would ensure fast, reliable inactivation of invasive DNA. However, it would also make sense to methylate cytosines slowly or with a certain lapse in time from the initial emergence of siRNAs. siRNAs also allow effective protection at the post-transcriptional level, so the postponed, non-impetuous decision to inheritably inactivate some genomic region by DNA methylation might be advantageous, because it could help to avoid potentially detrimental effects connected with unwanted permanent inactivation. Moreover, Pol V, which is regarded as an indispensable component of all RdDM pathways [3], was shown to be specifically attracted to methylated DNA [40], so the speed of the initial methylation of epigenetically naive *loci* could be restricted. Our results, based on the inducible activation of siRNA synthesis, show that RdDM could be initiated several hours after the appearance of high siRNA levels and that the targeted genomic region could reach practically full methylation in as early as 2 days in mitotically dividing tobacco BY-2 cells.

Results

The goal of our study was to describe the precise timing and progression of the transcriptional gene silencing (TGS) in its early stages. For this purpose, we used the BY-2 cell line [48] as a model, which allowed us to monitor the process in a highly synchronised and homogeneous culture, which can be studied at a single-cell level [49]. A selected BY-2 cell line stably expressing *GFP*, driven by *CaMV 35S* promoter (*P35S*) [50], was supertransformed with an estradiol-inducible silencing construct composed of an inverted repeat prepared from a part of *P35S* (*IR-P35S*) (Fig. 1a). The *IR-P35S* transcript was expected to form an RNA hairpin, which should have been processed to siRNAs targeting *P35S* and inducing TGS of the downstream laying *GFP* (Fig. 1b). For the experiment, we chose three independently transformed lines (8, 19, and 35), which showed a high fluorescence level, homogenous silencing response after *β*-estradiol treatment, and which did not show spontaneous self-silencing of the *GFP* gene. To keep the cells in a physiologically invariable state, cultures were continually kept in the exponential phase of growth. Establishment of TGS was monitored for 10 days of continuous *β*-estradiol treatment. In selected timepoints (3, 6, 12, 24 h and 2, 3, and 10 days), we determined the transcript level of the silencer and the target *GFP* gene, GFP fluorescence, promoter cytosine methylation, and presence of promoter-specific siRNAs in the selected lines.

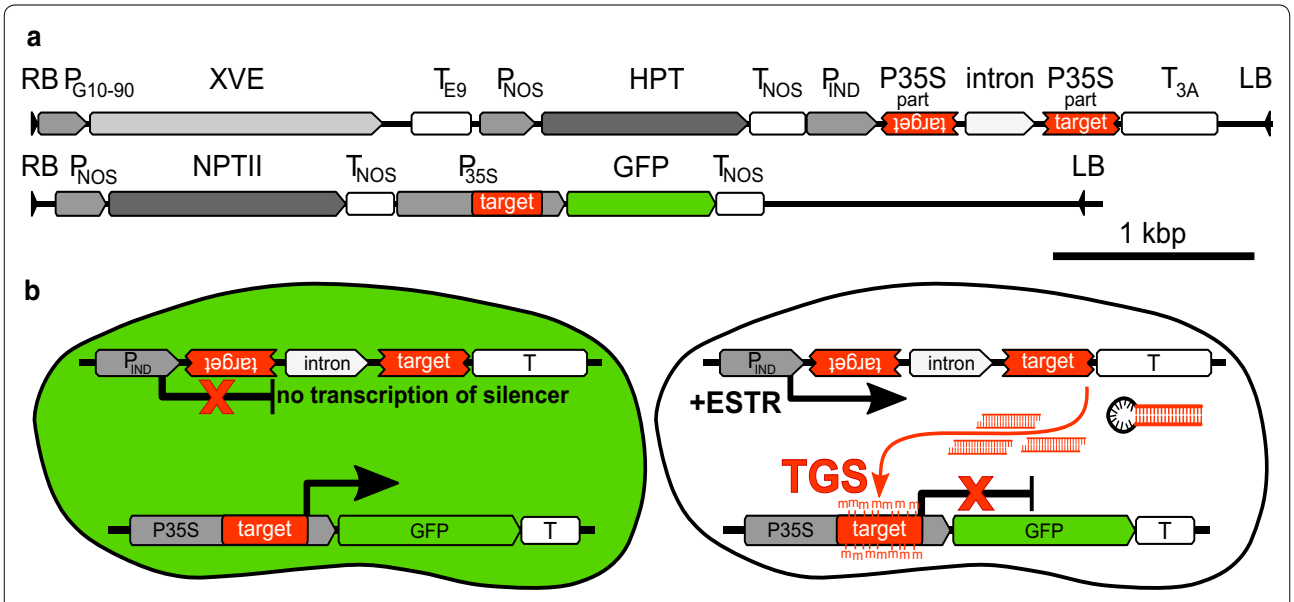


Fig. 1 Scheme of the model system. **a** Scheme of the inductor and the target *T-DNAs*; **b** expected activity of the T-DNAs in BY-2 cells untreated and treated with β-estradiol; P35S–CaMV constitutive promoter; XVE—a chimeric transcription activator; Pind—promoter activated by estradiol bound to XVE; HPT and NPTII, hygromycin, and kanamycin-resistance genes, respectively (for more detailed description, please see "Methods" and the original paper introducing the XVE-inducible system [70])

Monitoring of GFP silencing at the fluorescence and transcript level

GFP fluorescence after application of β-estradiol was assessed at a single-cell level using flow cytometry of isolated protoplasts. Ten-day β-estradiol treatment resulted in a complete loss of detectable GFP fluorescence in all three tested lines (Fig. 2a; Additional file 1). The fluorescence decreased quickly in lines 8 and 19, reaching about 50% of their initial level in 2 days, while the third line 35 responded more slowly (Fig. 2a). The flow-cytometry histograms clearly showed that in lines 8 and 19, the intensity of GFP fluorescence was highly homogeneous in the cell populations during the whole β-estradiol treatment (Additional file 1). It indicates that the cells responded synchronously to the induction, allowing a bulk analysis of harvested cells from these lines which provide reliable molecular data representing the progression of TGS in individual cells.

Following the results of the fluorescence analysis, we wondered how transcript levels of the silencer and the target *GFP* changed in the early steps of TGS. For this purpose, we analysed the homogeneously responding lines 8 and 19 using RT qPCR. After the application of β-estradiol, the transcript level of *IR-P35S* (estimated by amplification of the intron RNA indicating unspliced hairpin level) quickly increased within the first 3 h (Fig. 2b). Transcription of the silencer was followed by a rapid decrease in transcript level of the *GFP* gene in the first 12 h of the treatment ($p < 0.005$ for both lines) by about 30% (Fig. 2b). Afterwards, the transcription gradually declined to less than 5% of the initial transcript level within 2 days. The changes in GFP transcription were highly similar in both tested lines. *IR-P35S* transcript levels were, after the initial rise, fluctuating in time and finally decreasing towards the end of the treatment, even though the cells were continually exposed to β-estradiol, which might be connected with increased rate of the hairpin RNA cleavage by the action of not only DCLs, but also AGO proteins.

While the synthesis of new GFP protein was almost completely turned off during the first 2 days, the GFP fluorescence decreased by only 50% at the same time. Given that the cultures were exponentially growing, the observed decrease in GFP fluorescence had to result from both GFP degradation and GFP "dilution" in dividing cells. Comparing the speed of the GFP fluorescence decrease and the rate of BY-2 cell divisions (doubling time is about 20 h) [49] clearly indicated that the GFP protein was highly stable in BY-2 cells with a half-life of several days, which caused the observed delay in GFP fluorescence decline.

In summary, the fluorescence and transcription analyses clearly showed that the BY-2 cell populations homogeneously responded to β-estradiol and gradually switched off the GFP transcription during the first 2 days of the treatment.

The onset of *P35S* methylation

We further analysed how the observed decline in GFP expression (onset of TGS) correlated with methylation of the *P35S*. To analyse DNA methylation at a single nucleotide level, we used bisulfite modification of cytosines and subsequent sequencing of about 10 clones per sample. For the amplification, we designed primers, which covered not only the target, but also broader adjacent regions. Within the amplified segment, we obtained information about the methylation state of 90 cytosines in the target (379 nt) and 44 cytosines in the adjacent regions (104- and 82-nt up- and downstreams, respectively). From the 90 cytosines in the target region, there were 13 in CG and 9 in CHG context.

The analysis showed that the majority of analysed clones from lines 8 and 35 were practically without methylated cytosines (C*) (Fig. 2c; Additional file 2), and the frequency of non-methylated cytosines matched the experimentally determined efficiency of cytosine conversion, which was about 98% in our experiments. In contrast, in line 19, the median level of initial methylation was as high as about 11%, mainly due to the high proportion of symmetric CG methylation. This methylation might be a remnant of methylation induced by transient expression of the hairpin during the transformation event. Methylation in CHG and notably CHH context was at very low levels in all three lines before the treatment (Fig. 3).

After exposure to β-estradiol, the total cytosine methylation in lines 8 and 19 gradually increased and was significant after 12 h (Mann–Whitney *U* test, $p < 0.05$) (Fig. 2c). In the first 2 days, methylation reached its maximum of around 80% of C* in the target region. The methylation status of the target *P35S* region in lines 8 and 19 treated and untreated with β-estradiol was further confirmed in selected timepoints by cleavage with methylated DNA specific endonuclease (Additional file 3). The onset of methylation in line 35 was considerably slower, which was consistent with the later decline in GFP fluorescence in this line (Figs. 2a, c, 3). However, on day 10, the methylation reached similar levels in all three lines.

Methylation in CG and CHG contexts gradually increased in time and needed more than 3 days to reach their maximal levels. In contrast, CHH methylation showed a different pattern (Fig. 3). In lines 8 and 19, CHH methylation reached its maximum (median value near 80%) in 2 days and then slightly declined to about 60% until the 10th day. The median values of methylation in CG context on the 10th day of treatment reached 85%,

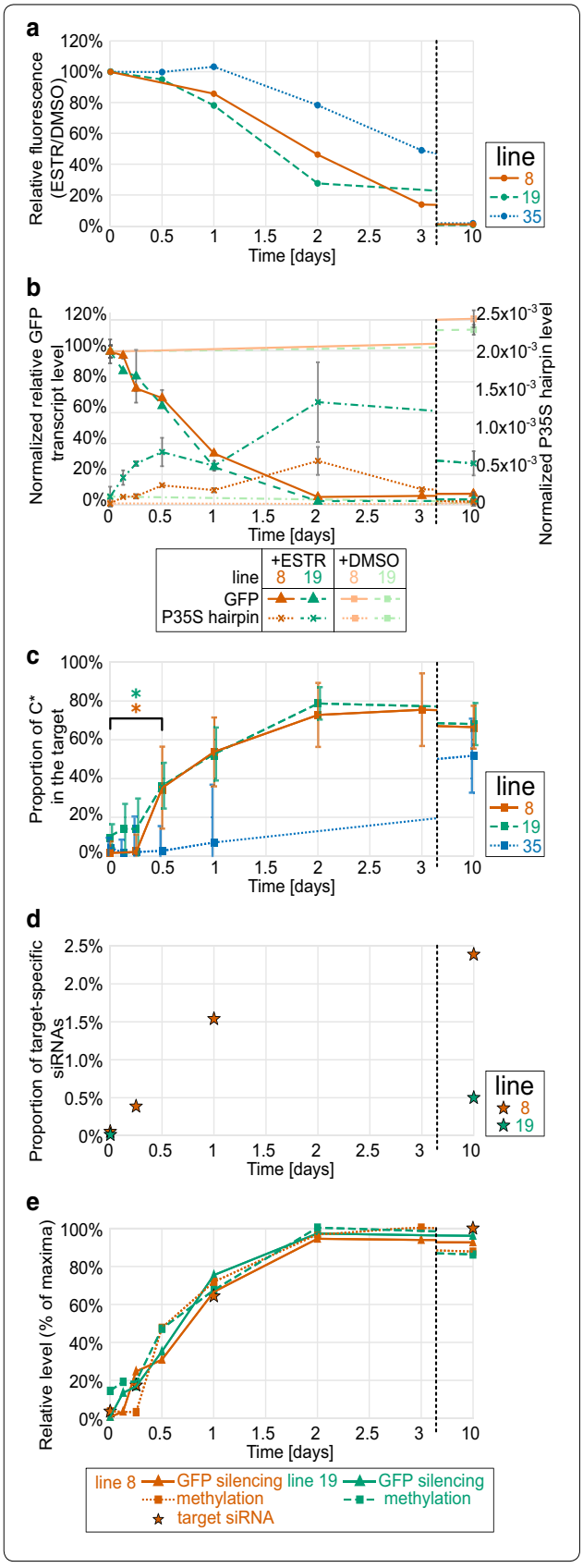


Fig. 2 Time course of transcriptional silencing of *P35S::GFP*. β-estradiol was applied at time 0 to three BY-2 lines (8, 19, and 35). The establishment of TGS of *P35S::GFP* was monitored in the selected timepoints at the level of **a** GFP fluorescence (flow cytometry of about 10 thousand cells per sample); **b** transcription of the *GFP* and the *P35S* promoter hairpin (RT qPCR); **c** proportion of methylated cytosines in the target region (means of about 10 clones per sample after bisulfite conversion); **d** proportion of promoter-specific siRNAs (of about 50 mil reads per sample); **e** correlation between the level of *GFP* silencing (i.e., relative decrease in *GFP* expression) and cytosine methylation (relative to the maximal attainable level) and siRNA level (relative to the maximal attainable value; only for line 8). Error bars in (**b**) and (**c**) indicate standard errors, and asterixis in (**c**) indicate a significant difference (Mann–Whitney *U* test, $p < 0.05$) for lines 8 and 19

92%, and 81% in lines 8, 19, and 35, respectively, whereas the methylation of CHG context was practically complete; the median reached 100%, 100%, and 94% in the respective lines. These results show that the dynamics of methylation establishment differed depending on the cytosine sequence context.

On the 10th day, methylated cytosines were more or less equally distributed along the target region. Even though there were regions and positions with less dense methylation, we could not find any cytosine position, which was utterly resistant to methylation in all analysed clones. The distribution of highly and less methylated cytosines along the target sequence was similar in all three analysed lines (Fig. 4). When focusing on the transient states, it seemed that there was a small preference to initiate methylation in the more upstream part of the target region (Additional file 2). The distribution of C* in incompletely methylated samples from earlier times indicates that the modification of some cytosines was "easier" and they reached their final methylation levels as early as after 24 h of the treatment, whereas other cytosines needed more extended time for the effective establishment of the methylation marks (Fig. 4).

In addition to the *P35S* region targeted by the *IR-P35S* construct, our methylation analysis also covered adjacent untargeted regions, which allowed us to evaluate the preciseness of targeting. Cytosine methylation was not absolutely restricted to the target region, but the proportion of methylated cytosines was much lower outside the target region. This external methylation was more prominent in the symmetrical contexts (mainly CHG) (Additional file 2) that was cumulated just upstream of the target region, which might be related to some pre-existing weak methylation present in this region before the treatment (Fig. 4, Additional file 2).

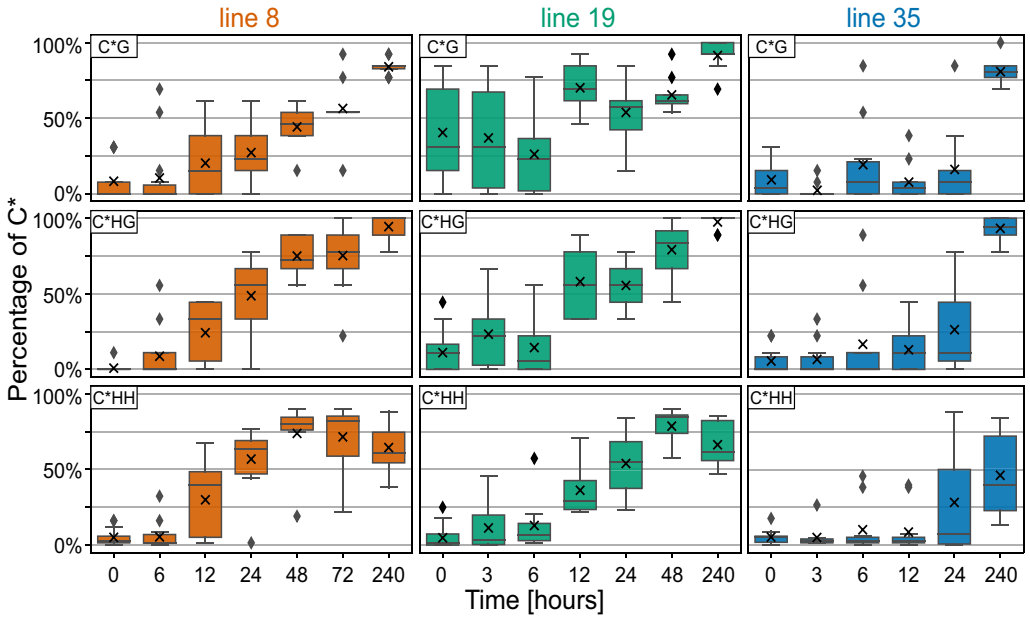


Fig. 3 Dynamics of cytosine methylation in different sequence contexts. Changes in cytosine methylation in CG, CHG, and CHH contexts in the target region of *P35S* were determined via bisulfite conversion in three BY-2 lines before and during β-estradiol treatment. The box plots show frequency of methylated cytosines in about 10 sequenced clones per sample. The analysis covered 68 cytosines in CHH, 13 in CG, and 9 in CHG context

In summary, the induction of the *P35S*-hairpin expression resulted in a gradual increase in *P35S* methylation within the first 2 days of exposure, reaching about 80% of methylated cytosines, which correlated well with the smooth attenuation of the *P35S* promoter activity (Fig. 2e).

Analysis of small RNAs as a trigger of promoter methylation

Sequence-specific DNA methylation is directed by siRNAs (RdDM) [25], so we performed high-throughput sequencing of siRNAs isolated from lines 8 and 19 at selected timepoints to see the correlation between the presence of promoter-specific siRNAs and promoter methylation. After checking the quality of sequencing output (Additional file 4), all siRNA reads were filtered for only those mapping on any region of the two T-DNAs present in our cells (containing the target *P35S::GFP* and the silencer *XVE::IR-P35S* regions; for the silencer, both spliced and unspliced forms were used for the filtering) (Additional files 5, 6: Table S2). The amount and distribution of siRNAs aligned to T-DNAs differed slightly between the two lines. Whereas in line 8, there were practically no siRNAs that aligned outside the target *P35S* region, in line 19 many regions of both T-DNAs were covered with low levels of siRNAs even before treatment (Additional files 5, 6: Table S2). After treatment with *β*-estradiol, siRNAs' levels increased mainly in the *P35S* target region, but in both tested lines, and more significantly in line 19, some smaller increases were also observed in the hygromycin phosphotransferase (*HPT*) expression cassette (composed of nopaline synthase promoter, *HPT* gene, and nopaline synthase terminator) lying upstream of the *P35S* hairpin (Additional files 5, 6: Table S2).

We further focused on the *IR-P35S*, where we analysed which sequences served as a source of siRNAs. As we expected, the alignment of siRNAs on the *IR-P35S* showed that the vast majority of siRNAs came from the inverted repeat region. No siRNAs aligned to the intron sequence and the exon/intron or intron/exon interface. Very low levels of siRNAs also originated from the spliced unique loop region of the hairpin, indicating the production of transitive secondary siRNAs (Additional file 5C), though these siRNAs could also originate from some structural rearrangements of the transgene [51]. The relative proportion of these transitive siRNAs did not significantly change during the treatment (Additional file 5C). In the target T-DNA containing full-length *P35S*, we detected no transitive siRNAs which would have expanded a single nucleotide from the target region at least in line 8 (Additional file 5).

For subsequent thorough analysis, we used only siRNAs aligning to the target *P35S*. In line 8, the level of these siRNAs gradually increased during the treatment, reaching almost 2.5% of all sequenced siRNAs on day 10. At the same time, in line 19, the level of *P35S* siRNAs was about 5 times lower (Fig. 2d). The length distribution of

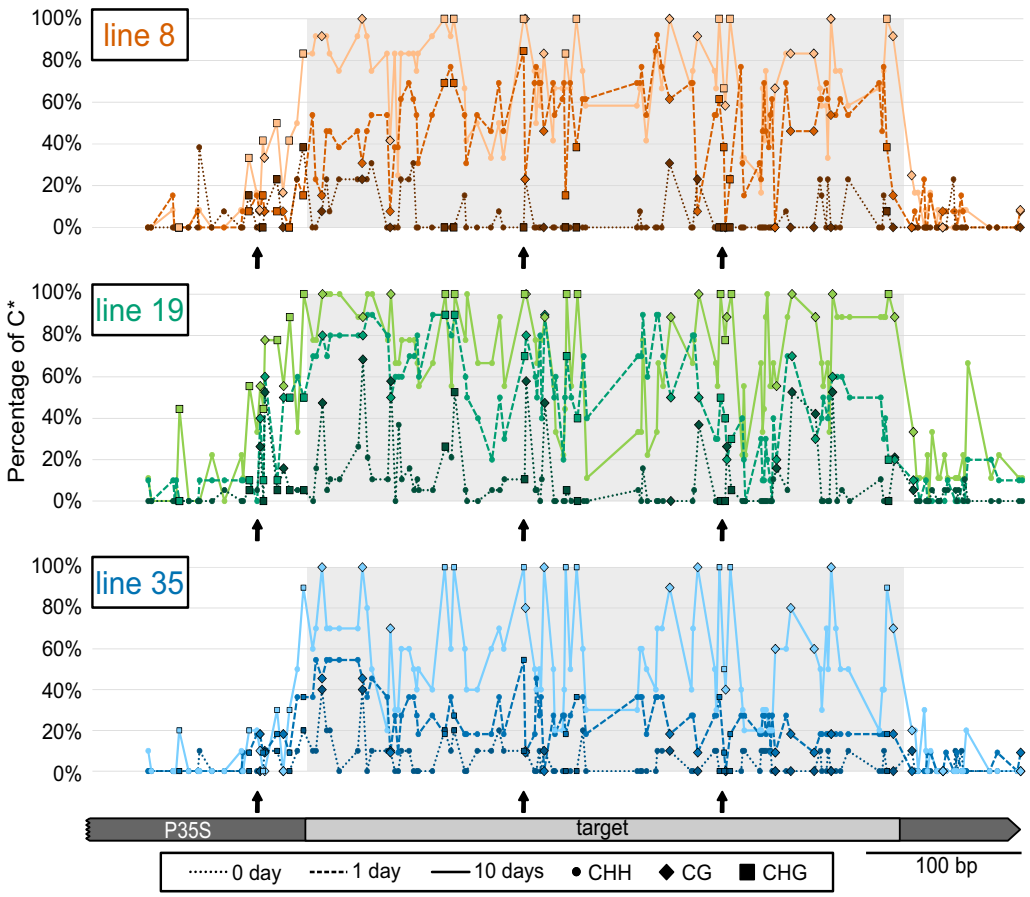


Fig. 4 Establishment of cytosine methylation and its distribution along the *P35S* region. The proportion of methylated cytosines on each cytosine position along the target and adjacent regions of *P35S* is shown for the three tested lines before the β-estradiol treatment and after 1 and 10 days of the treatment. Individual cytosine contexts are differentiated: circle for CHH, diamond for CG and square for CHG. Arrows indicate the position of CCG sites

siRNAs was relatively stable during the treatment, with 21-nt and 22-nt siRNAs being dominant (Fig. 5a). The ratio between 21 and 22-nt siRNAs differed in the two lines and fluctuated slightly during the treatment, but both size classes constantly formed together about 90% of all *P35S* siRNAs in both tested lines. 24-nt siRNAs stayed at an approximately constant level of about 5% of all *P35S* siRNAs during the whole treatment (Fig. 5b).

The siRNAs aligning to the *P35S* were not distributed equally along the sequence, but there were hot- and cold spots with high and low coverage by siRNAs, indicating different stability and/or efficiency in their generation (see the list of the most frequent siRNAs in Additional file 7: Table S3). The distribution of the hot- and cold spots was strongly strand specific. In most positions along the target region, the siRNAs aligned almost exclusively on either the forward or the reverse strand, and rather exceptionally to both of them (Fig. 5d). Despite this, the ratio between forward and reverse siRNAs stayed approximately constant during the treatment at about 1:1 in both tested lines (Fig. 5c). The position of hot- and cold spots was similar for all siRNA size categories. Whereas the distribution of 21- and 22-nt siRNAs was almost identical, 24-nt siRNAs in line 8 aligned to slightly different positions, especially in the most upstream part of the target region (Fig. 5e). The distribution of siRNAs along the sequence was also relatively stable during the whole treatment, though some gradual changes could be seen with increasing duration of the *β*-estradiol treatment; especially in the case of 24-nt siRNAs aligning to the left border of the target region in line 8 (Fig. 5f). Interestingly, low levels of *P35S* siRNAs were already detected in the untreated cells. The size and strand-specific distribution of these sRNAs along the sequence corresponded with the situation observed in the induced cultures, indicating that these siRNAs likely originated from a few sporadically silenced cells, which can also be seen in the flow-cytometry histograms (Additional file 1).

Surprisingly, the distribution of siRNAs along the target region differed between the two analysed lines. Though the position of the main hot- and cold spots was

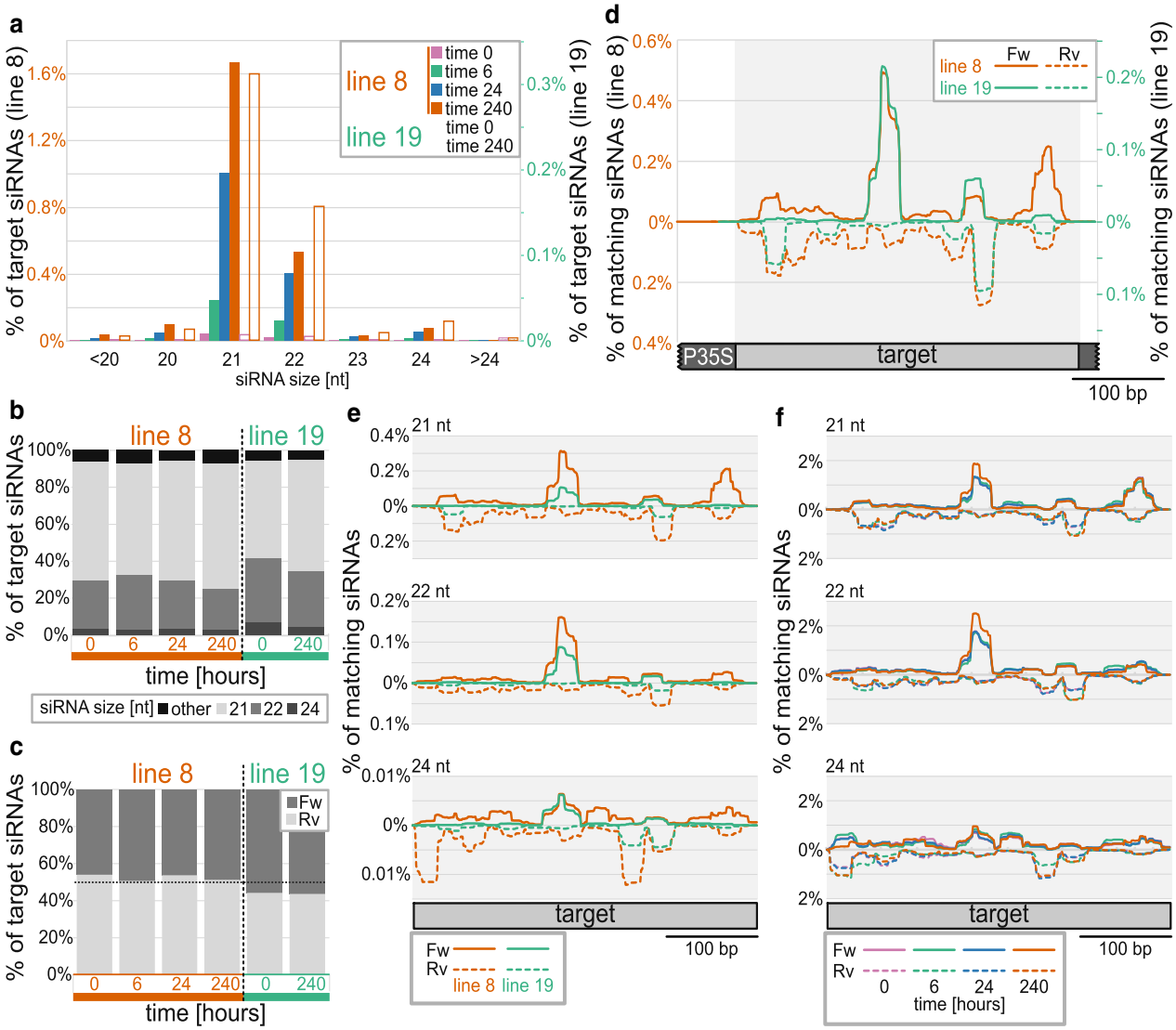


Fig. 5 Classification and distribution of siRNAs matching with the *P35S* region. All sequenced siRNAs were filtered for those matching with the *P35S* target region and further analysed: **a** relative representation of size classes of *P35S* siRNAs in all sequenced siRNAs (1% corresponds to 1×10^4 siRNAs per million reads; note different scales for the two lines); **b** relative representation of selected siRNA size classes in *P35S*-specific siRNAs; **c** relative representation of forward- and reverse-oriented *P35S*-specific siRNAs; **d** distribution of siRNAs isolated from 10-day treated cultures along the promoter sequence depicted as % of all sequenced siRNAs (note different scales for the two lines); **e** distribution of selected siRNA size classes along the promoter sequence depicted as % of all sequenced siRNAs (siRNAs isolated from 10-day treated cultures); **f** time changes in distribution of selected siRNA size classes along the promoter sequence in line 8 (depicted as % of siRNAs of the respective size matching to the target in the respective time)

similar, in the case of line 8, the coverage by siRNAs was much more homogeneous. In line 19, there were sharper peaks and also deeper valleys, with only a minimal number of aligning siRNAs (Fig. 5d). The regions with low siRNA coverage in line 19 and higher coverage in line 8 included approx. 30-nt-long border part of the target region. However, there was no obvious difference in the distribution of final methylation along the target region in the two lines on day 10 (Additional file 2, Fig. 4).

In summary, the transcription of *IR-P35S* led to the formation of high levels of target-specific siRNAs with the dominant representation of 21- and 22-nt-long classes. The siRNAs aligned unevenly and with strand-specificity along with the target *P35S* region, causing relatively smooth and homogeneous methylation of the whole target region. The levels of target-specific siRNAs gradually increased during the treatment, but there were no dramatic changes in either the representation of siRNA size categories or their distribution along the target region.

Discussion

Our study describes in detail the dynamics of de novo RdDM in plant cells. Thanks to our highly synchronised BY-2 model system, we could show that induction of

DNA methylation and subsequent transcriptional silencing can be a rapid process. Methylation was initiated almost immediately with the appearance of promoter-specific siRNAs, and the first significant difference in methylation was observed as soon as 12 h after induction and caused complete transcriptional silencing within 2 days. Later, however, the methylation pattern developed further. There were some more unexpected features that we described in connection with the establishment of TGS; CG methylation occurring at the target locus prior to the treatment had no effect on the speed of RdDM, the proportion of 24-nt siRNAs was low and stable in the 1st days of DNA methylation, the speed of de novo methylation was much slower than maintenance methylation of newly synthesised DNA strands, and the dynamics of DNA methylation differed depending on the cytosine sequence context. Our study thus opens up new questions that will lead to a better understanding of this biologically important process.

High levels of specific siRNAs quickly induce methylation

Expression of the *P35S* hairpin led to the gradual accumulation of siRNAs to relatively high levels as compared with the previous reports [52]. *P35S* siRNAs exceeded levels of siRNAs aligning onto well-characterised tobacco transposable elements by a factor of 10 (Tnt1 and Tto1) (Additional file 6: Table S2) [53, 54]. The total levels of *P35S* specific siRNAs differed about five times in the two tested lines, with no effect on the speed of methylation (Fig. 2), indicating that the siRNAs levels were above saturation, so we likely observed the maximal speed of de novo DNA methylation.

The siRNA level sufficient for full, dense methylation in line 19 on day 10 corresponded to the siRNA level observed in line 8 as early as after 6 h of the treatment. However, at that time, only weak methylation was present in line 8 (Fig. 2). It clearly indicates that not only reaching a certain level of siRNAs, but also a certain duration of the exposure to siRNAs was necessary to establish dense methylation of the target region. On the other hand, once established, the dense methylation had to be very quickly introduced to newly synthetized DNA strands after replication, because we observed this dense methylation in the cells that were continually dividing approximately every 20 h [49]. Such faster methylation compared to de novo methylation of a naive locus was likely connected with the presence of the maternal highly methylated DNA strand, since Pol V is effectively recruited to *loci* with methylated cytosines [40].

The onset of dense cytosine methylation was gradual, and the effectiveness of methylation varied slightly depending on the cytosine position within the target region (Fig. 4). However, we did not detect any clear correlation with the position of hot and cold spots of either typical 21–24-nt siRNA (Fig. 5) or longer sRNAs (practically missing in our system) (Additional file 6: Table S2), which could potentially represent sidRNAs that were suggested to participate in the initiation of methylation in epigenetically naive *loci* [44]. The onset of methylation also differed, depending on the cytosine sequence contexts. Methylation in CG and CHG contexts reached higher final levels compared to CHH methylation, which likely reflected the fact that we monitored methylation in dividing cells. CG and CHG methylation could be reintroduced to newly synthesised strands more rapidly and more infallibly after the replication than CHH, because once established, CG and CHG could also be methylated independently of RdDM [18, 55]. Recently, it was shown that external cytosines of CCG sites (a subtype of CHG) in gene bodies could only be methylated when internal cytosines are methylated [56]. Only two such sites present in our target sequence, however, showed the opposite tendency. In line 18 lacking initial CG methylation, the methylation of the external cytosine preceded the methylation of the internal one, which was probably associated with the combined action of DRM2 and maintenance methylation by CMT3. The decrease in CHH methylation observed during prolonged 10-day treatment might result from silencing of the IR construct and a hypothetical decrease in siRNA levels between day 2 (not determined in our siRNA analysis) and day 10. However, the decrease in CHH methylation was equal in both tested lines, though the level of siRNAs on day 10 was 5 times higher in line 8 compared to line 19. Therefore, we prefer an alternative explanation: establishment of high-density CG and CHG methylation might reduce the necessity or efficiency of RdDM, which is then somehow attenuated irrespective the continual presence of high levels of siRNAs.

It should also be noted that the proportion of 24-nt siRNAs was relatively low and stable, as was also reported for IR-derived siRNAs in *Arabidopsis* [52]. This indicates that either the observed levels of 24-nt siRNAs were sufficient to induce effective methylation or that other siRNA sizes participated or were fully responsible for targeting de novo methylation in our system. Involvement of 21-nt and 22-nt siRNAs was clearly demonstrated in the RDR6–AGO6 RdDM pathway [57]. Since there was no relative increase in 24-nt siRNA production during the treatment, it is also unlikely that Pol IV was involved in the generation of siRNAs from this locus, even when it was already repressed and densely methylated on day 10. All detected siRNAs more likely originated from the Pol II hairpin transcript. The high frequency of 22-nt siRNAs indicated the involvement of DCL2, which was shown to stimulate the synthesis of secondary siRNAs through

RDR6 activity [58]. Therefore, the detected siRNAs were likely a mixture of primary siRNAs produced from the stem part of the inducer hairpin and secondary siRNAs, presumably originating from the inducer transcript processed by RDR6 after the primary AGO cleavage [59].

This assumption was further supported by the detection of siRNAs originating from outside the dsRNA (stem) region of the hairpin, although levels of these siRNAs were relatively low compared to those originating from the IR region (Additional file 5C). On the other hand, transitivity was completely missing in the target locus. It can be linked to the fact that we targeted a promoter sequence, which is not transcribed or is transcribed only sporadically [60], so the transcripts could not serve as a source of secondary siRNAs. The absence of siRNAs from outside the target region was consistent with practically no CHH methylation outside the target region, showing high preciseness of RdDM targeting, as was suggested from the molecular mechanism of DRM2 action [25]. Such pinpoint targeting is especially important for the inactivation of invasive DNAs (transposable elements) inserted between plant genes, whose expression should not be affected [13].

In contrast to CHH, the methylation of CHG was also relatively high in the adjacent regions and reached practically 100% within the target region, indicating the involvement of CMT3 with less precise targeting, which is based on CMT3 binding to H3K9me2 [12]. This chromatin mark is known to attract not only CMT3, but also Pol IV (via SHH1) [31]. However, our results did not indicate the production of 24-nt siRNAs from *P35S* by Pol IV/RDR2/DCL3 activities, as we see no increase in the proportion of 24-nt siRNA in later times. Therefore, either the Pol IV occupancy was prevented by another chromatin/histone modification (e.g., H3K4me3) [31], or the presence of CHG methylation (the activity of CMT3) is not necessarily connected with H3K9me2. This situation has already been documented in tobacco by ChIP, where high CHG methylation in transcriptionally silenced *P35S* was unexpectedly accompanied by H3K9 acetylation and H3K4me3 marks [61]. H3K4me3 activation marks coexisted with methylated cytosines also in many human promoters, whose activity was frequently unaffected by the introduction of methylation marks [62].

In our system, the levels of GFP silencing strongly correlated with *P35S* methylation (Fig. 2e), indicating that promoter activity could be smoothly modulated by gradually increasing methylation levels. Recently, it was demonstrated that the effect of DNA methylation in P35S depended on the position of methylcytosines, including strand affiliation [63]. In our case, methylation was induced on long P35S region, so there was no step change in the GFP transcription connected with methylation of specific cytosine or passing over a hypothetical threshold methylation level, but instead, an even regulation was possible along with the wide range of methylation levels.

Methylation was induced even on DNA free of methylcytosines

It has long been known that the presence of sRNAs can effectively trigger de novo RdDM of euchromatic *loci*. However, Pol V, which is considered to be involved in the final step of all canonical and non-canonical RdDM pathways [3, 43], should be attracted exclusively to methylated DNA through its interaction with SRA-domain proteins SUVH2/9 [40, 64]. Since DNA methylation is not commonly present in euchromatic DNA, the precise mechanism of de novo methylation of epigenetically naive *loci* remains unclear. In our experiments, the target and adjacent promoter regions were practically free of methylated cytosines just before treatment in two out of the three tested lines (lines 8 and 35), but in both these lines, the expression of *P35S* hairpin triggered de novo cytosine methylation. Moreover, the progression and speed of de novo methylation in line 8 were fully comparable with line 19 characterised by a relatively high level of initial CG methylation. Since methylated CG should be specifically recognised by SUVH2 [64], the impact of pre-existing cytosine methylation for Pol V activity or the role of Pol V in de novo DNA methylation remains disputable.

The independence of Pol V activity from the pre-existing methylation was recently documented in *suvh2/9* mutants, which were not significantly impaired in de novo methylation of *LTR* (long-terminal repeat) from an exogenous TE introduced into the *Arabidopsis* genome [43]. Whereas SUVH2/9 might be omitted, the authors showed that Pol V was indispensable for methylation of newly introduced TEs in T1 plants. On the contrary, it was recently shown that Pol V was not needed for the methylation of viral DNA [65], indicating the existence of an alternative pathway responsible for the establishment of DNA methylation, at least on certain occasions. In invertebrates and yeasts, which lack specialised RdDM polymerases, Pol II is supposed to serve as an enzyme assisting in the targeting of RdDM or RNA-directed histone modifications [40]. Therefore, based on the current knowledge, the involvement of Pol II in the initial targeting of de novo methylation cannot be excluded even in plants, whereas Pol V remains indispensable for a specific recognition of TE (LTR) sequences and later efficient maintenance RdDM via the canonical pathway.

Conclusions

Methylation of cytosines in the promoter region is known to down-regulate expression of a downstream gene. In our study, we analysed the timing of transcriptional

silencing of the *GFP* gene in proliferating tobacco BY-2 cell lines, which provided a highly synchronised and homogeneous response. This model enabled us to demonstrate, to our knowledge for the first time, that the induction of DNA methylation and subsequent transcriptional silencing can be a rapid process, initiated practically immediately with the appearance of promoter-specific siRNAs. These siRNAs were mostly 21- and 22-nt long and gradually accumulated at very high levels, forming up to 2.5% of all detected siRNAs on day 10. Relative distribution of siRNAs strongly differed along the promoter sequence with no transitivity observed in the target region. Our data also indicated that CG methylation occurring in the target region before the treatment did not affect the speed of RdDM. The dynamics of DNA methylation differed depending on the cytosine sequence context with gradually increasing methylation in the symmetrical CG and CHG contexts during the whole 10-day treatment, while CHH methylation reached its maxima already after 2 days and then slightly decreased. During the 2-day exposure to siRNAs, which was sufficient for the establishment of dense methylation in the target region, a gradual increase in the proportion of methylated cytosines smoothly attenuated promoter activity.

Methods

Plant materials

The *Nicotiana tabacum L.* cell line BY-2 [48] was cultivated in a medium based on the Murashige and Skoog (MS) [66] formula; MS salts (Merck) were supplemented with 200 mg/L K_2HPO_4, 100 mg/L myo-Inositol, 3% sucrose, 1 mg/L vitamin B1, and 1 μM 2,4-D, pH adjusted to 5.8 with 1 M KOH. Cell cultures were cultivated at 27 °C in darkness in 100 mL Erlenmeyer flasks on an orbital shaker at 110 rpm. The cell lines were subcultured weekly by 0.7 mL into 30 mL fresh media, and continually exponential cultures were subcultured every 3–4 days by 1.5 mL. Non-homogeneous cultures (in respect of GFP expression) were subcloned [50] before starting the experiments. A BY-2 line carrying smRS-GFP (called simply GFP in the paper) [67] stably expressed under the control of constitutive *CaMV 35S* promoter (*P35S*) for many years (Fig. 1a) [50]. The line was super-transformed with a hairpin construct prepared from PCR amplified 379-bp-long segments of the *P35S* arranged as a head-to-head inverted repeat, separated by an intron originating from the *PsbO1* gene of *Solanum tuberosum* (*PUT-157a-Solanum_tuberosum-62673150*) [68] with short adjacent regions (Fig. 1a) [69]. Expression of the hairpin was controlled by the *β*-estradiol (Sigma) *XVE*-inducible system [70]. Selected super-transformed lines were treated by adding *β*-estradiol to a final concentration of 2 μM (from 20 mM stock solution in DMSO stored at −20 °C) into the cultivation media; controls were treated with a corresponding concentration of DMSO.

Fluorescence analysis

For fluorescence analysis, protoplasts were prepared by taking 1.5 mL of the cell culture into a 2 mL tube, and the medium was drained off with cellulose wadding tampons. 1.5 mL of a protoplast enzyme mixture (10 g/L Cellulase, 1 g/L Pectolyase Y-23 in 0.45 M D-mannitol) was added, and the whole mixture was transferred into a 6-well cell-culture plate and incubated in the dark for 3 h at 26 °C with shaking 90 rpm on an orbital shaker. Protoplasts were sedimented in a 2 mL tube at 200 RCF for 5 min, and the pellet was resuspended with 1 mL of MS with 0.4 M sucrose. Protoplasts were floated by centrifugation (200 RCF for 5 min) without braking. 200 μL of the upper phase was used for flow-cytometry analysis using BD LSR II. Measured particles were first gated to select live protoplasts (Additional file 1A) and analysed using FlowJo vX.0.7 (https://www.flowjo.com/).

Transcription analysis

Transcript levels of the *GFP* and the *P35S* hairpin were analysed by RT qPCR. RNA was isolated from 100 mg of biomass using the RNeasy® Plant Mini Kit (QIAGEN) and reverse transcribed with RevertAid Reverse Transcriptase (Thermo Fisher Scientific) using anchored oligoT_{23} primer. Quantification of the *GFP* and the *P35S* hairpin transcript levels was performed on a LightCycler 480 (Roche) using the iQ TM SYBR Green Supermix (BioRad, Hercules, USA) with primers, as listed in Additional file 8: Table S1. All reactions were performed in triplicate. The PCR product specificity was verified by melting curve analysis using a LightCycler 480 software. The PCR efficiency and Cq values were calculated using the software LinRegPCR 2017.1 [71]. Calculated concentrations were normalised to the expression of the internal expression standard *EF1α* [72] with primers adopted for tobacco *EF1α* genes [73].

DNA-methylation analysis

DNA-methylation analysis was performed using bisulfite conversion with the EpiTect Bisulfite Kit (QIAGEN) as described previously [74]. Primers for amplification of *P35S* region (Additional file 8: Table S1) were designed to anneal on the converted DNA, irrespective of its original methylation state. About ten cloned PCR products for each sample were sequenced and analysed using the MS Excel 2016 and Python 3. The level of methylation was further confirmed by qPCR after cleavage of genomic DNA with McrBC endonuclease (New England Biolabs)

specific for methylated DNA (modified from [75]). In brief, 100 ng of DNA was first fragmented with a restriction enzyme that does not cut it the region of interest (AseI), and then, the reaction was split into half and supplemented with 10 units of McrBC enzyme or equivalent amount of 50% (v/v) glycerol. The DNA was digested for 6 h at 37 °C, and then, the enzyme was inactivated (20 min, 65 °C). 1 ng was used for qPCR performed as described in the section "Transcription analysis" with primers listed in Additional file 8: Table S1.

Small RNA analysis

The RNA samples were isolated from 100 mg (FW) of BY-2 cells with the RNeasy Plant Mini Kit (QIAGEN). RNA was quality assessed and quantified. A fraction of sRNAs ranging in size from 18 to 45 nt were excised and recovered from 15% urea–polyacrylamide gels. Extracted sRNAs were ligated with 5′ and 3′ RNA adapters with T4 RNA ligase. The adapter-ligated small RNAs were subsequently transcribed into cDNA by Super-Script II Reverse Transcriptase (Invitrogen) and amplified using adaptor-specific primers. The amplified cDNA products were size-purified and circularised (ssDNA circles). This sRNA library was sequenced using the combinatorial probe–anchor synthesis (cPAS)-based BGISEQ-500 sequencer provided at an affordable price (BGI, Shenzhen, China), which was previously shown to provide highly reproducible results comparable with other NGS platforms [76]. Obtained raw data were analysed in the software Geneious 11.1.5 (https://www.geneious.com) and MS Excel 2016; only perfectly matching siRNAs were used for analyses. The sRNA data sets used in this study are available in the following database European Nucleotide Archive PRJEB32154 (http://www.ebi.ac.uk/ena/data/view/PRJEB32154).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13072-019-0299-0.

Additional file 1. Analysis of GFP fluorescence in BY-2 Protoplasts.

Additional file 2. Distribution of methylated cytosines in the P35S region.

Additional file 3. Estimation of the P35S methylation by McrBC cleavage.

Additional file 4. Quality scores for sRNA sequencing.

Additional file 5. Distribution of siRNAs along the two T-DNAs before and after 10-day treatment with β-estradiol.

Additional file 6: Table S2. Characterization of sequenced siRNAs (siRNA numbers per 1 million reads).

Additional file 7: Table S3. List of the most frequent siRNAs matching with the *P35S* target (numbers of detected siRNAs per 1 million reads).

Additional file 8: Table S1. List of primers used in the study.

Acknowledgements
We thank Šárka Motylová for providing the *IR-P35S* construct and Lena Hunt for language corrections.

Authors' contributions
AP performed the vast majority of experimental work, processed and analysed data and participated in the study design and writing the manuscript. VC performed some methylation and qPCR analyses and participated in the study design. DT performed some methylation and qPCR analyses. LF designed the study and together with AP wrote the manuscript. All authors read and approved the final manuscript.

Funding
The work was supported by the Czech Ministry of Education, Youth and Sports [LO1417].

Availability of data and materials
The sRNA data sets used in this study are available in the following database European Nucleotide Archive PRJEB32154 (http://www.ebi.ac.uk/ena/data/view/PRJEB32154).

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 4 June 2019 Accepted: 22 August 2019
Published online: 11 September 2019

References

1. Bennetzen JL. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev. 2005;15:621–7.
2. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16:519–32.
3. Zhang H, Lang Z, Zhu J-K. Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol. 2018;19:489.
4. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The *Arabidopsis thaliana* mobilome and its impact at the species level. eLife. 2016;5:e15716.
5. Zhou S, Liu X, Zhou C, Zhou Q, Zhao Y, Li G, et al. Cooperation between the H3K27me3 chromatin mark and non-CG methylation in epigenetic regulation. Plant Physiol. 2016;172:1131–41.
6. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell. 2008;135:635–48.
7. Fojtová M, Houdt HV, Depicker A, Kovarik A. Epigenetic switch from post-transcriptional to transcriptional silencing is correlated with promoter hypermethylation. Plant Physiol. 2003;133:1240–50.
8. Saze H, Kakutani T. Differentiation of epigenetic modifications between transposons and genes. Curr Opin Plant Biol. 2011;14:81–7.
9. Finnegan EJ, Peacock WJ, Dennis ES. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. PNAS. 1996;93:8449–54.
10. Jones L, Ratcliff F, Baulcombe DC. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Curr Biol. 2001;11:747–57.
11. Woo HR, Pontes O, Pikaard CS, Richards EJ. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Gene Dev. 2007;21:267–77.
12. Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012;151:167–80.

13. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205.
14. Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002;416:556–60.
15. Malagnac F, Bartee L, Bender J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. 2002;21:6842–52.
16. Ebbs ML, Bartee L, Bender J. H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. Mol Cell Biol. 2005;25:10507–15.
17. Ebbs ML, Bender J. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell. 2006;18:1166–76.
18. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007;17:379–84.
19. Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell. 2014;55:495–504.
20. Stoddard CI, Feng S, Campbell MG, Liu W, Wang H, Zhong X, et al. A nucleosome bridging mechanism for activation of a maintenance DNA methyltransferase. Mol Cell. 2019;73:73–83.e6.
21. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001;292:2077–80.
22. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell. 2013;152:352–64.
23. Bologna NG, Voinnet O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. Annu Rev Plant Biol. 2014;65:473–503.
24. Elvira-Matelot E, Martínez ÁE. Diversity of RNA silencing pathways in plants. Plant Gene Silenc Mech Appl. 2017;5:1–31.
25. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell. 2014;157:1050–60.
26. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005;308:118–20.
27. Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell. 2005;120:613–22.
28. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, et al. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet. 2005;37:761–5.
29. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi M-A, et al. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. Gene Dev. 2005;19:2030–40.
30. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet. 2011;7:e1002195.
31. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase-IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498:385–9.
32. Zhang H, Ma Z-Y, Zeng L, Tanaka K, Zhang C-J, Ma J, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. PNAS. 2013;110:8290–5.
33. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, et al. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. eLife. 2015;4:e09591.
34. Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, et al. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. Genome Res. 2015;25:235–45.
35. Zhai J, Bischof S, Wang H, Feng S, Lee T, Teng C, et al. A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. Cell. 2015;163:445–55.
36. Zhang Z, Liu X, Guo X, Wang X-J, Zhang X. *Arabidopsis* AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing. Nat Plants. 2016;2:16049.
37. El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, et al. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Gene Dev. 2007;21:2539–44.
38. Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, et al. Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. Cell Res. 2016;26:66–82.
39. Kuhlmann M, Mette MF. Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in *Arabidopsis thaliana*. Plant Mol Biol. 2012;79:623–33.
40. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014;507:124–8.
41. Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. 2014;10:e1003948.
42. Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet. 2013;45:1029–39.
43. Fultz D, Slotkin RK. Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. Plant Cell. 2017;29:360–76.
44. Ye R, Chen Z, Lian B, Rowley MJ, Xia N, Chai J, et al. A Dicer-independent route for biogenesis of siRNAs that direct DNA methylation in *Arabidopsis*. Mol Cell. 2016;61:222–35.
45. Vaucheret H, Institut N de la RA. Promoter-dependent trans-inactivation in transgenic tobacco plants: kinetic aspects of gene silencing and gene reactivation. C R Acad Sci III. 1994:310–23.
46. Philips JG, Dudley KJ, Waterhouse PM, Hellens RP. The rapid methylation of T-DNAs upon agrobacterium inoculation in plant leaves. Front Plant Sci. 2019. https://doi.org/10.3389/fpls.2019.00312.
47. Xu C, Corces VG. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. Science. 2018;359:1166–70.
48. Nagata T, Nemoto Y, Hasezawa S. Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. In: Jeon KW, Friedlander M, editors. International review of cytology. Cambridge: Academic Press; 1992. p. 1–30. https://doi.org/10.1016/s0074-7696(08)62452-3.
49. Srba M, Černíková A, Opatrný Z, Fischer L. Practical guidelines for the characterization of tobacco BY-2 cell lines. Biol Plant. 2016;60:13–24.
50. Nocarova E, Fischer L. Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. BMC Plant Biol. 2009;9:44.
51. Jupe F, Rivkin AC, Michael TP, Zander M, Motley ST, Sandoval JP, et al. The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genet. 2019;15:e1007819.
52. Wroblewski T, Matvienko M, Piskurewicz U, Xu H, Martineau B, Wong J, et al. Distinctive profiles of small RNA couple inverted repeat-induced post-transcriptional gene silencing with endogenous RNA silencing pathways in *Arabidopsis*. RNA. 2014;20:1987–99.
53. Grandbastien M-A, Spielmann A, Caboche M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature. 1989;337:376.
54. Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. PNAS. 1996;93:7783–8.
55. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, et al. *Arabidopsis* MET1 cytosine methyltransferase mutants. Genetics. 2003;163:1109–22.
56. Zabet NR, Catoni M, Prischi F, Paszkowski J. Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies. Nucleic Acids Res. 2017;45:3777–84.
57. Panda K, Ji L, Neumann DA, Daron J, Schmitz RJ, Slotkin RK. Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. Genome Biol. 2016;17:170.
58. Parent J-S, Bouteiller N, Elmayan T, Vaucheret H. Respective contributions of *Arabidopsis* DCL2 and DCL4 to RNA silencing. Plant J. 2015;81:223–32.

59. Sijen T, Vijn I, Rebocho A, van Blokland R, Roelofs D, Mol JN, et al. Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr Biol. 2001;11:436–40.
60. Čermák V, Fischer L. Pervasive read-through transcription of T-DNAs is frequent in tobacco BY-2 cells and can effectively induce silencing. BMC Plant Biol. 2018;18:252.
61. Křížová K, Depicker A, Kovařík A. Epigenetic switches of tobacco transgenes associate with transient redistribution of histone marks in callus culture. Epigenetics. 2013;8:666–76.
62. Ford E, Grimmer MR, Stolzenburg S, Bogdanovic O, Mendoza A de, Farnham PJ, et al. Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation. 2017. http://bioRxiv.org/170506.
63. Matsunaga W, Shimura H, Shirakawa S, Isoda R, Inukai T, Matsumura T, et al. Transcriptional silencing of 35S driven-transgene is differentially determined depending on promoter methylation heterogeneity at specific cytosines in both plus- and minus-sense strands. BMC Plant Biol. 2019;19:24.
64. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet. 2008;4:e1000280.
65. Jackel JN, Storer J, Coursey T, Bisaro D. *Arabidopsis* RNA polymerases IV and V are required to establish H3K9 methylation, but not cytosine methylation, on geminivirus chromatin. J Virol. 2016;90:7529–40.
66. Murashige T, Skoog F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant. 1962;15:473–97.
67. Davis SJ, Vierstra RD. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol. 1998;36:521–8.
68. Duchoslav M, Fischer L. Parallel subfunctionalisation of PsbO protein isoforms in angiosperms revealed by phylogenetic analysis and mapping of sequence variability onto protein structure. BMC Plant Biol. 2015;15:133.
69. Motylová Š. The influence of RDR6 activity and mode of RNAi induction on dynamics and mechanism of silencing of the reporter GFP gene in tobacco cell line BY-2. Diploma thesis. Faculty of Science, Charles University. 2015. http://hdl.handle.net/20.500.11956/74403.
70. Zuo J, Niu Q-W, Chua N-H. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. 2000;24:265–73.
71. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. 2009;37:e45.
72. Nicot N, Hausman J-F, Hoffmann L, Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot. 2005;56:2907–14.
73. Dvořáková L, Srba M, Opatrny Z, Fischer L. Hybrid proline-rich proteins: novel players in plant cell elongation? Ann Bot. 2012;109:453–62.
74. Tyč D, Nocarová E, Sikorová L, Fischer L. 5-Azacytidine mediated reactivation of silenced transgenes in potato (*Solanum tuberosum*) at the whole plant level. Plant Cell Rep. 2017;36:1311–22.
75. Bond DM, Baulcombe DC. Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in *Arabidopsis thaliana*. PNAS. 2015;112:917–22.
76. Fehlmann T, Reinheimer S, Geng C, Su X, Drmanac S, Alexeev A, et al. cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. Clin Epigenetics. 2016;8:123.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BBA - Gene Regulatory Mechanisms 1863 (2020) 194647

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Unexpected variations in posttranscriptional gene silencing induced by differentially produced dsRNAs in tobacco cells

Vojtěch Čermák, Dimitrij Tyč, Adéla Přibylová, Lukáš Fischer *

Charles University, Faculty of Science, Department of Experimental Plant Biology, Viničná 5, Prague 2 128 44, Czech Republic

ARTICLE INFO

Keywords:
DNA methylation
dsRNA
PTGS
RNAi
siRNA
Tobacco BY-2 cell line

ABSTRACT

In plants, posttranscriptional gene silencing (PTGS) is induced by small RNAs (sRNAs) generated from various dsRNA precursors. To assess the impact of dsRNA origin, we compared downregulation of GFP expression triggered by inverted repeat (IR), antisense (AS) and unterminated sense (UT) transcripts transiently expressed from the estradiol-inducible promoter. The use of homogeneously responding tobacco BY-2 cell lines allowed monitoring the onset of silencing and its reversibility. In this system, IR induced the strongest and fastest silencing accompanied by dense DNA methylation. At low induction, silencing in individual cells was binary (either strong or missing), suggesting that a certain threshold sRNA level had to be exceeded. The AS variant specifically showed a deviated sRNA-strand ratio shifted in favor of antisense orientation. In AS lines and weakly induced IR lines, only the silencer DNA was methylated, but the same target GFP sequence was not, showing that DNA methylation accompanying PTGS was influenced both by the level and origin of sRNAs, and possibly also by the epigenetic state of the locus. UT silencing appeared to be the least effective and resembled classical sense PTGS. The best responding UT lines behaved relatively heterogeneously possibly due to complexly arranged T-DNA insertions. Unlike IR and AS variants that fully restored GFP expression upon removal of the inducer, only partial reactivation was observed in some UT lines. Our results pointed out several not yet described phenomena and differences between the long-known silencer variants that may direct further research and affect selection of proper silencer variants for specific applications.

1. Introduction

RNA interference (RNAi) is an important mechanism involved in the regulation of gene expression in eukaryotic cells. It is often used in research as a tool to downregulate expression of a selected gene. The key players in RNAi are small RNAs (sRNAs), which can be formed via several different pathways in plants, making plant RNAi a very complex phenomenon [1]. Generally, a double stranded RNA (dsRNA) is needed to induce the production of sRNAs in plants. These dsRNAs are recognized by DICER-LIKE (DCL) proteins, which process them into sRNA duplexes. One strand of the sRNA duplex (guide strand) is then selected to associate with ARGONAUTE (AGO) proteins, while the other strand, so-called passenger strand, is degraded. AGO with a loaded guide strand recognizes various target RNA molecules by sequence complementarity. Four DCL genes and ten AGO genes are encoded in the *Arabidopsis thaliana* genome. The DCLs differ according to what type of dsRNA they prefer to dice and the length of the sRNAs they produce [2–5]. The AGO proteins bind sRNAs based on their length and the identity of their 5′ nt and select the guide strand by a thermodynamic mechanism in cooperation with DRB proteins [6–10]. AGO proteins can slice the RNA target or cooperate with other proteins to direct various processes like blocking translation or directing DNA methylation (which involves a wide range of proteins e.g. SUO or DRM2, respectively) [11,12]. When the target is an mRNA molecule, then these processes lead to posttranscriptional gene silencing (PTGS). However, the same pool of sRNAs can also induce chromatin modifications (primarily DNA methylation), which can accompany the PTGS [13]. Such methylation does not significantly affect gene transcription, because it is usually limited to the transcribed region [14]. When this methylation spreads to the promoter region, then the PTGS can change to silencing at the transcriptional level (TGS; [15]).

In plants, DNA can be methylated at cytosines in both symmetrical (CG and CHG) and asymmetrical (CHH) contexts, which principally differ in the mechanism of their maintenance in newly synthesized strands after replication [16,17]. Hemimethylated CG sites can be easily recognized and maintained by METHYLTRANSFERASE 1 (MET1) in cooperation with VIM proteins [18]. However, additional signals from

* Corresponding author.
E-mail address: lukasf@natur.cuni.cz (L. Fischer).

https://doi.org/10.1016/j.bbagrm.2020.194647
Received 22 April 2020; Received in revised form 15 October 2020; Accepted 19 October 2020
Available online 27 October 2020
1874-9399/© 2020 Elsevier B.V. All rights reserved.

histones are needed for non—CG (CHG and CHH) methylation maintenance. DNA methylation is mutually linked to histone modifications through chromatin modifying enzymes, whose activity is either directly or indirectly controlled by the other type of the chromatin mark. Thus, CHG and CHH methylation is maintained in heterochromatin by CHROMOMETHYLASE 2 (CMT2) and CHROMOMETHYLASE 3 (CMT3) attracted by histone modification (H3K9me2) recognized directly by chromodomains of these enzymes [19,20]. Vice versa, SRA domains of H3K9 histone methyltransferases such as KRYPTONITE (KYP, SUVH4) and SUVH5/6 recognize DNA methylation [21–23]. The other signal for maintenance methylation is presence of hc-siRNAs that are involved in the canonical RNA-directed DNA methylation (RdDM) pathway. These hc-siRNAs are generated from products of RNA polymerase IV that transcribes chromatin with specifically modified histone H3 (H3K4me0, H3K9me2) which is recognized by SHH1 [24]. DNA methylation in RdDM is introduced by DOMAINS REARRANGED METHYLASE 1/2 (DRM1/2) when specific AGO protein carrying sRNA binds to complementary nascent scaffold transcripts generated by plant-specific RNA polymerase V (Pol V) specifically attracted to methylated DNA by SUVH2/9 [12,25]. In addition to this maintenance methylation, RdDM also establishes de novo DNA methylation utilizing sRNAs of various origin including those arising during PTGS [13,26].

There are three basic ways how dsRNA can be formed. First, separate transcription of complementary (sense and antisense) RNA strands leads to their intermolecular pairing, generating dsRNA. This occurs naturally in the pathway generating natural-antisense siRNA (nat-siRNAs) [27]. The second way is based on the transcription of an inverted repeat; such a transcript creates a hairpin dsRNA through intermolecular pairing. Naturally, miRNAs and some other types of hairpin siRNAs are generated in this way [4,28]. The third way uses RNA-DEPENDENT RNA POLYMERASE (RDR) to convert certain ssRNAs into dsRNAs. This mechanism is naturally involved in the production of hc-siRNAs, ta-siRNAs and most of the secondary siRNAs [2,29,30]. It is also involved in the sense transgene silencing (S-PTGS) and can be connected with the phenomenon known as cosuppression [31] that is believed to result from high production of aberrant RNAs. These RNAs can saturate the RNA degradation pathway leading to their accumulation in the cell, which makes them available as a substrate for RDRs [32].

Many types of silencer constructs have been designed based on this knowledge and have been meticulously compared in an attempt to reach the highest level of silencing and the maximum percentage of silenced transformants (e.g. [13,33–35]). We hypothesized that the use of constructs based on different ways dsRNA forms would lead to the activation of different RNAi pathways and thus results in a different course of silencing, even if lines with similar decrease in expression are selected. The understanding of all the effects dsRNA formation has on the silencing outcomes is not only important for understanding the RNAi itself, but it is also important for the use of appropriate silencing constructs in applied research.

In our study, we chose the three most distinct ways dsRNA forms and prepared the following silencer constructs: antisense (AS), unterminated (UT; expected to produce aberrant RNAs; [32]) and inverted repeat (IR); all of them based on (and targeting) the full-length coding sequence of *GREEN FLUORESCENT PROTEIN* (*GFP*) gene. We used an inducible promoter for these constructs to eliminate spontaneously silenced lines and to be able to observe the silencing from the very beginning to the steady state and during its termination (recovery of GFP expression). We supertransformed each of these constructs to the highly homogeneous tobacco BY-2 cell line [36] where they targeted *GFP* transcripts stably expressed from the identical locus in all the variants [37–39]. This allowed us to make direct comparisons of the effects of the three silencer variants. Although the model system based on cell lines and silencing of the *GFP* transgene was quite artificial, it allowed to control most of the variables so the observed differences could be attributed directly to the way dsRNA formed. With a similar system, we recently described the dynamics of the initial phases of de novo RdDM [39]. In this current work, we first compared dynamics of PTGS onset at the fluorescence level in populations of hundreds of independently transformed lines (calli). Thereafter, the silencing dynamics and recovery at mRNA, sRNA and DNA methylation levels and by flow cytometry at the single cell fluorescence level were analyzed in detail in selected lines with the highest silencing rate.

2. Results

2.1. Establishment of lines for inducible silencing and their comparisons at the population level

Three different silencer constructs were prepared to induce GFP silencing (Fig. 1). All were based on the full-length coding sequence of *GFP* gene and their transcription was controlled by β-estradiol inducible promoter (XVE system; [40]). The constructs were as follows: antisense *GFP* (AS), inverted-repeat *GFP* (IR) and unterminated *GFP* (UT) where the sense transcripts were expected to be aberrantly terminated and recognized by RDR (RNA-dependent RNA polymerase) to produce dsRNA. As a control in some experiments, we also used a sense *GFP* gene with a terminator (GF) that can under certain occasions undergo sense transgene posttranscriptional silencing (S-PTGS) connected with cosuppression of the target GFP. In cases where the *GFP* sequence was in the sense orientation (UT and GF variant), it was without the start codon (unless otherwise stated) to prevent potential unwanted translation of the GFP from the silencer RNA. Also, no additional start codon was present in the sequence that could lead to the production of a truncated yet fluorescent protein.

We transformed these constructs into the BY-2 cell line with stable expression of GFP [37]. Surprisingly, about 70% of transformants carrying the IR construct showed spontaneous silencing without induction with estradiol [38], so only minority of IR lines could be used for further experiments.

From the sets of independently transformed lines carrying the individual silencer constructs, we chose those that showed sufficiently homogeneous and high GFP fluorescence (about one hundred for each variant) and planted one half of the callus on control medium (D) and the other half on induction medium with estradiol (E) and measured the fluorescence of those calli during 14 day period (Table 1; Fig. S1A, B; Table S2). From these data, we analyzed the efficiency (percentage of lines able to induce silencing), strength (the maximum level of decreased expression) and speed of silencing. The IR was fastest in silencing induction and was able to induce the strongest silencing in a larger number of lines than any other variant. The ability of AS to induce silencing was somewhat lower and the UT was the least effective. Surprisingly, the UT silencing did not significantly differ from the cosuppression caused by the GF variant, which was used as a control to see the effect of the presence/absence of terminator on the silencing efficiency (Table 1). Both UT and GF variants contained only very few effectively silenced lines. Although the construct without terminator should produce aberrant RNAs, it was not always the case. Surprisingly, about 50% of lines, where an UT silencer construct with the start codon was transformed into the wild type BY-2 cells, were able to produce some GFP fluorescence (Fig. S1C). This indicates that in these lines the RNA was at least partially polyadenylated possibly due to T-DNA insertion in front of an endogenous terminator or due to a specific T-DNA arrangement. Otherwise, the mRNA could not be exported from the nucleus and translated [41].

Percentage of lines with inducible silencer constructs or with the control sense *GFP* (GF) that were able to induce strong silencing of GFP expression (reduction of GFP fluorescence below 50% of mock-induced control) at different time points after the transfer to β–estradiol containing medium (lower number of available IR calli was caused by the high frequency of spontaneous silencing). Statistics for this data are in Table S2 and biological replicates are shown in Fig. S1A.

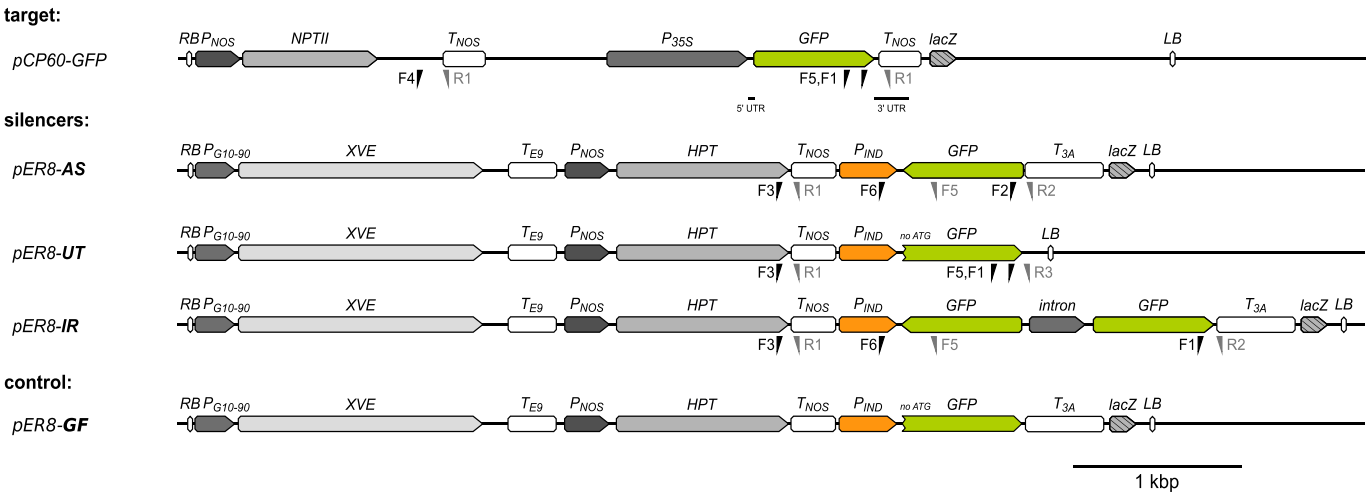


Fig. 1. Scheme of the silencer and target *T-DNAs*.
Functional elements and positions of primers used for reverse transcription, qPCR and McrBC assay are shown at the appropriate sequences (primers itself are listed in Table S1). Target *T-DNA* of *pCP60-GFP* (https://www.addgene.org/122173) confers constitutive GFP expression. *pER8*: a set of *T-DNAs* used for the inducible expression: *pER8-AS*: *GFP* gene in antisense orientation; *pER8-UT*: sense *GFP* gene without terminator; *pER8-IR*: sequence of *GFP* arranged as an inverted repeat; *pER8-GF*: sense *GFP* with terminator as a control. In the UT and GF constructs, the GFP sequence was without start codon, unless stated otherwise. *RB*: right border, *LB*: left border, P_{NOS}: nopaline synthase promoter, P_{35S}: *35S* promoter, P_{G10-90}: synthetic constitutive promoter; P_{IND}: inducible promoter activated by β-estradiol; T_{NOS}: nopaline synthase terminator; T_{E9}: *rbcS E9* terminator; T_{3A}: *rbcs S 3A* terminator; *XVE*: estrogen receptor and transcriptional activator; *NPTII*: kanamycin resistance gene; *HPT*: hygromycin resistance gene; *GFP*: green fluorescent protein coding sequence; intron: intron from *Solanum tuberosum PsbO* gene; *LacZ*: fragment of bacterial β–galactosidase gene. All *T-DNAs* are at the same scale as indicated with 1 kbp scale bar.
Functional elements and positions of primers used for reverse transcription, qPCR and McrBC assay are shown at the appropriate sequences (primers itself are listed in Table S1). Target *T-DNA* of *pCP60-GFP* (https://www.addgene.org/122173) confers constitutive GFP expression. *pER8*: a set of *T-DNAs* used for the inducible expression: *pER8-AS*: *GFP* gene in antisense orientation; *pER8-UT*: sense *GFP* gene without terminator; *pER8-IR*: sequence of *GFP* arranged as an inverted repeat; *pER8-GF*: sense *GFP* with terminator as a control. In the UT and GF constructs, the GFP sequence was without start codon, unless stated otherwise. *RB*: right border, *LB*: left border, P_{NOS}: nopaline synthase promoter, P_{35S}: *35S* promoter, P_{G10-90}: synthetic constitutive promoter; P_{IND}: inducible promoter activated by β-estradiol; T_{NOS}: nopaline synthase terminator; T_{E9}: *rbcS E9* terminator; T_{3A}: *rbcs S 3A* terminator; *XVE*: estrogen receptor and transcriptional activator; *NPTII*: kanamycin resistance gene; *HPT*: hygromycin resistance gene; *GFP*: green fluorescent protein coding sequence; intron: intron from *Solanum tuberosum PsbO* gene; *LacZ*: fragment of bacterial β–galactosidase gene. All *T-DNAs* are at the same scale as indicated with 1 kbp scale bar.

Table 1
Efficiency of GFP silencing in BY-2 calli.

Variant	% of silenced lines			# of analyzed lines
	0 days	3 days	10 days	
AS	0.0	0.9	11.6	112
UT	0.0	2.0	9.0	100
GF	1.0	1.0	9.1	99
IR	1.6	30.6	45.2	62

2.2. Detailed analyses of the time course of silencing

For further experiments, we selected three lines of each variant (named IR1-3, AS1-3 and UT1-3) with highly homogeneous *GFP* fluorescence (determined by flow cytometry at the single cell level) and with the highest silencing rate, as such lines are preferentially used in reverse genetics. Comparing lines with the potential to fully induce silencing also allowed us to look for specific differences between the silencer variants that were not directly dependent on the differences in silencing strength between the variants.

The estradiol inducible system enables to adjust the strength of silencer transcription by changing the concentration of estradiol in the

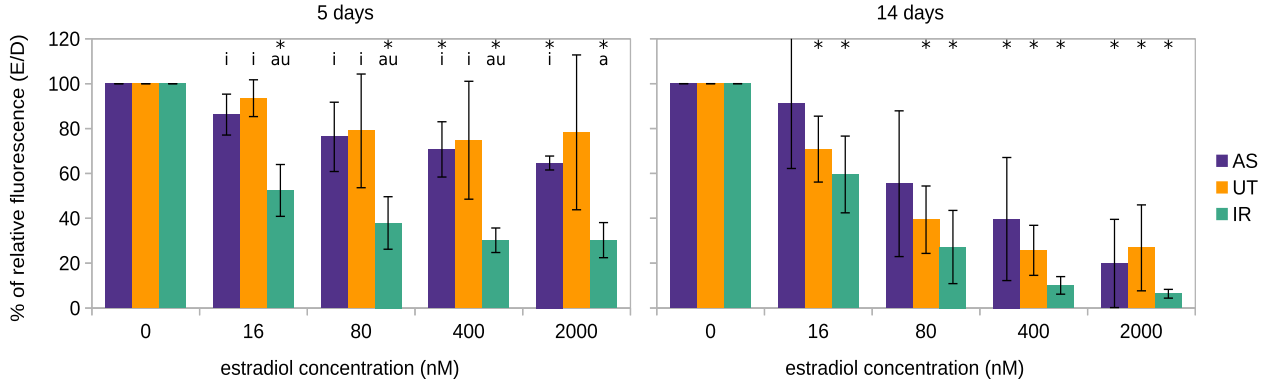


Fig. 2. Effects of estradiol concentration on the induction of silencing.
The ratio of relative GFP fluorescence between calli cultured on medium with indicated estradiol concentrations and without estradiol (E/D) for 5 and 14 days. Data were normalized for day 0 to be 100%. For each silencer (AS, UT, IR) three lines were tested and for each line three calli were measured. The error bars represent standard deviation. The asterisk indicate statistically significant difference ($p < 0.05$) from 0 nM estradiol treatment of the same silencer variant; the letters "a", "i" and "u" indicate statistically significant difference ($p < 0.05$) from the AS, IR and UT variant treated with the same estradiol concentration.

medium. We first tested the GFP silencing in calli of the selected lines on 2 μM concentration (commonly used in the XVE system) and a series of dilutions (Fig. 2). The IR showed the fastest and strongest response, even on 16 nM concentration of estradiol. For the IR variant, the effect of the estradiol concentration was most prominent at the final state of silencing, rather than at the speed of silencing. Interestingly, the UT silencer was also able to induce silencing even on 16 nM estradiol when given enough time, and the differences in the strength of silencing were the least dependent on the estradiol concentration in this variant. In contrast, the AS variant showed a strong dependence on the estradiol concentration, with 16 nM concentration being too low to induce any significant silencing (Fig. 2).

The previous experiments suggested that the standard estradiol concentration (2 μM) was saturating for most lines, so it was used in the subsequent experiments. The three selected lines for each variant were transferred to suspension cultures and exposed for 14 days to estradiol and then the estradiol was washed out and the cells were kept for additional 21 days on the standard medium. Control cells for each line were handled in the same way, but without exposure to estradiol.

Based on the fluorescence measurements using a flow cytometer (Figs. 3A; S2A, D) silencing in the IR lines was already obvious on the first day and by the 14th day, the fluorescence decreased to 1–11% of the original fluorescence level, reacting almost completely homogeneously. After removal of estradiol, the fluorescence mostly returned to the original level. The AS lines showed the first signs of a reaction later, on the second day, and by the 14th day decreased to 6–20% of the original level. This reaction was mostly homogeneous and fluorescence returned almost to the original levels after removal of the inducer. The reaction of UT lines was more heterogeneous compared to other variants with the UT1 line resembling IR lines (Fig. S2A). On the 14th day, the fluorescence decreased to 6–30% of the original level. After removal of the inducer, many cells of UT3 and UT2 line were able to maintain the GFP silencing (Fig. S2D).

Changes in RNA transcript levels were faster than changes in fluorescence (Figs. 3B; S2B), probably due to the long lifetime of the GFP protein [39]. At the transcript level, all IR lines reacted within the first 6 h after the exposure to estradiol ($p < 0.05$, paired t-test), reaching 20–74% of the original transcript level (Fig. S2B). The silencing reached its maximum as early as in 1–2 days (2–11%). Only one AS line and one UT reacted so rapidly within 6 h. The responses of the other lines were delayed; the first decreases were detected after one day. Silencing in AS and UT lines progressively grew during the whole treatment and reached 12–36% and 17–34% after 14 days for the AS and UT lines, respectively.

Surprisingly, transcript levels of the silencer were highest between 6 and 24 h for AS and UT but peaked between 24 and 48 h for IR (Figs. 3C; S2C). Moreover, the IR silencer transcript levels were about 10-fold lower than those for AS and UT (with the exception of the UT1 line).

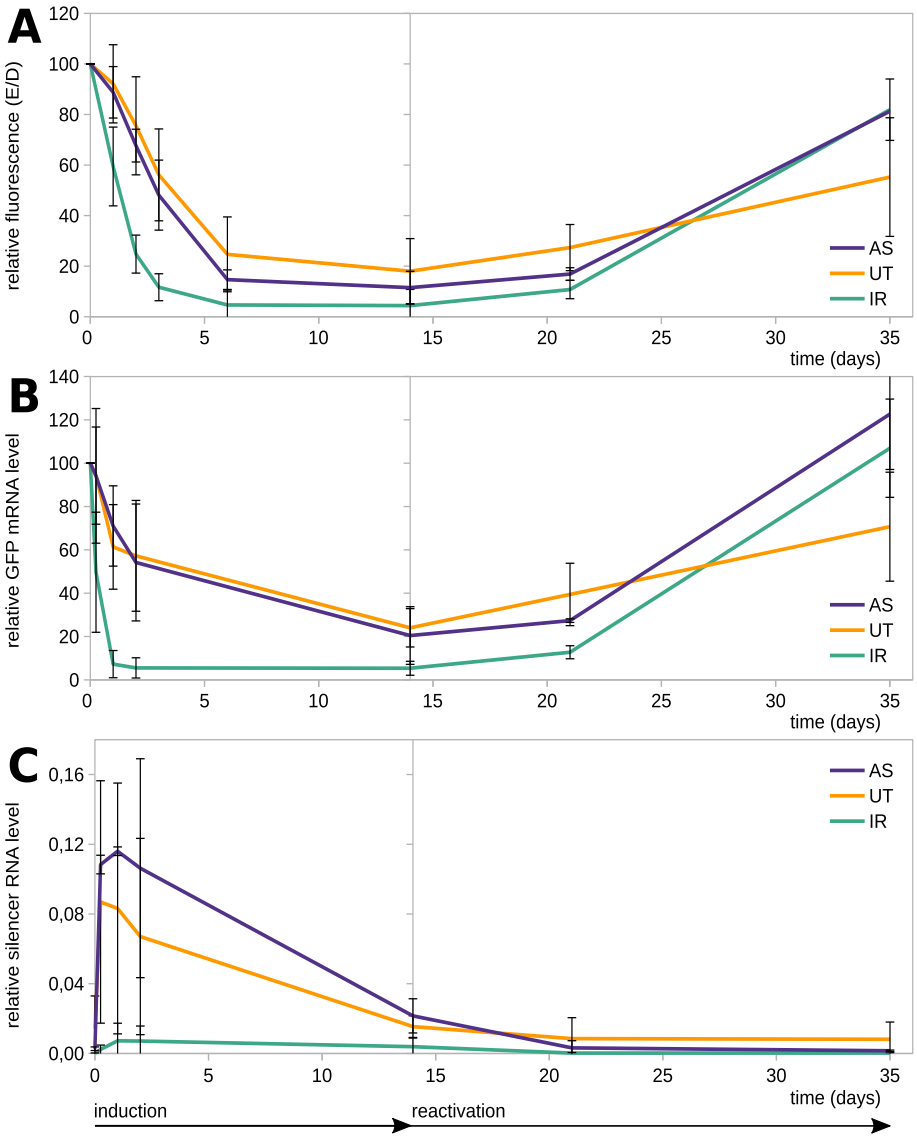


Fig. 3. Dynamics of GFP silencing and recovery in suspension cultures.
Three best responding lines for each silencer variant were treated with β–estradiol for 14 days, then β–estradiol was washed out and the cells were monitored for an additional 21 days.
(A) Time course of GFP fluorescence as measured by flow cytometer. The relative fluorescence is shown as the ratio of fluorescence between induced and mock-induced cells (E/D). Data were normalized for time 0 to be 100%. Data for AS2 on the 35th day are not available.
(B) Time course of *GFP* transcript levels as measured by RT-qPCR. The relative *GFP* transcript levels were normalized to the internal standard *EF1α* and related to time 0 that was set to 100%.
(C) Time course of silencer transcript levels as measured by RT-qPCR. The relative silencer transcript levels were normalized to *EF1α*.
Error bars in all plots represent standard deviation. Data for each line individually are shown in Fig. S2.
Error bars in all plots represent standard deviation. Data for each line individually are shown in Fig. S2.

2.3. IR silencing under weak induction

The IR variant showed stronger response than the other variants, so we analyzed its course of silencing in more detail on estradiol concentrations near the expected threshold for silencing induction: 0, 4, 16 and 64 nM. Flow cytometry data of the average fluorescence and RNA expression data showed that two IR lines were sensitive enough to induce silencing, even on 4 nM estradiol, whereas the IR3 line began to react on 16 nM estradiol (Fig. S3A, B). This indicates that these concentrations were really at the lowest edge for silencing induction. At these concentrations, the silencing in individual cells (seen on flow cytometry data in Figs. 4, S3C) was not homogeneous (continuous in the strength of response), but was present in two distinct states – active and silenced.

2.4. Molecular characterization of the best responding UT lines

The behavior of the UT1 line compared to the UT2 and UT3 was abnormal and resembled silencing induced by IR (stronger and faster onset, almost complete reversion of silencing after removal of estradiol, and very low silencer transcript levels). It led us to examine the arrangement and copy number of *T-DNAs* in the three UT lines, because UT *T-DNAs*, when inserted as tail-to-tail inverted repeats (LB to LB), might produce a hairpin RNA. The number and size of hybridizing restriction fragments on Southern blot (a higher number of fragments detected by *GFP* probe after *Kpn*I digestion; Fig. S4) indeed suggest the presence of such an inverted repeat in the UT1 line, though we were unable to confirm its presence by PCR (data not shown). But it is known that PCR over inverted repeats is problematic [42]. Southern blot analysis did not indicate the presence of inverted repeats in UT2 and UT3 lines, but it suggested the presence of direct-tandem repeat in the UT3 line and indeed we were able to confirm it by PCR (Fig. S4E) and sequence the *T-DNA* junction (Fig. S4F).

The larger number of detected *T-DNA* insertions raised a question of the possibility of read-through transcripts at the *T-DNAs*, either from other *T-DNAs* in case where they would be in the repeat arrangement or from genomic DNA, resembling results from our previous study [38]. Indeed, we were able to detect low levels of both sense and antisense RNAs over different regions of the silencer in all three UT lines (Fig. S4G, H). Some of these RNAs were present already before induction, while others appeared in the response to estradiol induction. This behavior could be explained by the T-DNA arrangement as a result of readthrough transcripts from neighboring T-DNAs (in UT1 and UT3 lines) or by the activity of RDRs, as antisense RNA levels from genic regions behaved differently compared to intergenic regions. From highly expressed genic regions targeted by sRNAs, RDRs were shown to synthesize detectable amounts of complementary RNAs [30].

2.4.1. Analyses of small RNAs

The key determinant of PTGS is sRNAs, so we sequenced their populations in lines with a typical course of silencing for IR and AS variants (IR1 and AS3). Due to heterogeneity in UT lines, we could not select a really typical line, so we decided to analyze the line UT3, which was interesting in that it remained strongly silenced even after estradiol removal. Samples for this analysis were collected during the experiment analyzing the time course of silencing at the three key time points in the experiment (0, 14 and 35 days; see Fig. 3). Sequenced sRNAs were mapped on the target and the silencer *T-DNAs*. Some low levels of sRNAs aligning to the target sequence were detected even before induction (especially in the case of IR; Fig. 5A) which might have originated from a weak spontaneous transcription of the silencer [38], although it seemed that they had no effect on silencing. Estradiol induction led to the accumulation of target-specific sRNAs in all variants in quantities reaching 10 to 100 times the original level. The target T-DNA-specific sRNAs accounted for 0.5 to 1.9% of all sRNAs sequenced from the cells (Figs. 5A; S5A, B; Table S3A). After removal of estradiol, the high levels of sRNAs detected in IR declined, even beyond the initial levels. In contrast, target-specific sRNAs did not decrease in the UT3 line, but even increased further (Fig. 5A), which could explain the maintenance of silencing observed in UT lines after estradiol removal (Figs. 3A, B; S2A, B; data for AS at day 35 are not available).

2.4.1.1. Target T-DNA-specific sRNAs. At the target *T-DNA*, almost all sRNAs matched with the *GFP* transcription unit (more than 91.4%; Fig. S5A). In the case of IR, the sRNAs almost exclusively matched with only the *GFP* coding sequence, whereas in the case of AS and UT, there was strong 3′ transitivity to the terminator *3′UTR* region (Fig. 5A, B). No obvious 5′ transitivity at the target *T-DNA* was seen in any of the samples (Fig. S5C).

The silencer variants also differed in the representation of sRNAs of different lengths, more specifically, in the ratios of 21/22 nt sRNAs (Figs. 5C; S5E). The AS and UT had a dominant fraction of 21 nt sRNAs (49 and 40% respectively), while 22 nt sRNAs dominated in the IR (49%). The levels of 24 nt sRNAs were similar among the variants (around 20%) and the other sRNA sizes made up only a small fraction (less than 8% altogether).

The ratio of sRNAs aligning to the forward and reverse strand of the target *GFP* gene was around 50% for IR and UT variants, but the AS variant deviated, with levels of the reverse strand sRNAs twice as high compared to forward strand sRNAs (Fig. 5D; 5: Fig. S5E). The variants did not show any larger differences in the representation of the 5′ nucleotide of the target specific sRNAs (Fig. S5F).

sRNAs mapping to the target *GFP* transcription unit were not distributed equally (Figs. 5B, E; S5G), but they formed distinct peaks on either the reverse or forward strand. These peaks had highly similar positions between different sRNA size classes. Positions of peaks in the same region were also similar between different variants, but the variants differed in the heights of individual peaks and their representation along the sequence. In the IR variant, most of the sRNAs came from the 5′ part of the *GFP* gene; in the AS, the distribution of peaks was more or less homogeneous along the whole transcription unit (including *3′UTR*); and in the UT, most of the sRNAs came from the 3′ end – actually 41% of sRNAs in UT at the GFP transcription unit come from the *3′UTR*, which was not primarily targeted by the silencer. In the AS line, high production of secondary sRNAs can also be inferred thanks to a point mutation in the silencer GFP sequence introduced during cloning (A to T at 404. nucleotide from the start of GFP cds, see Fig. S5H for the resulting AS silencer sequence). This single point mutation was expected to have no influence on the silencing, but it could be used to partially distinguish between primary and secondary sRNAs; sRNAs where the antisense strand lacks the mutation or where the sense strand has the mutation cannot be primary sRNAs. Considering this criterion, at least 60% of sRNAs overlapping this nucleotide position had to be secondary (Table S3B).

2.4.1.2. sRNAs specific to the silencer T-DNAs. In the case of the silencers, there were more regions outside the primary sRNA source sequence that showed accumulation of sRNAs after the estradiol induction (Fig. S5B). As with the target *GFP*, the AS and UT silencers showed 3′ transitivity (although weaker), but no obvious 5′ transitivity (Fig. S5C). Interestingly in the UT silencer, which has no terminator, the levels of sRNAs sharply decreased at the end of the *GFP* sequence and showed smaller accumulation of sRNAs at the region downstream of the *GFP* coding sequence than in the terminated AS silencer (Fig. S5B, C).

Beside sRNAs matching the silencer transcripts, there were also sRNAs from the *HPT* (hygromycin phosphotransferase) transcripts in the AS and UT variants (Fig. S5B, C, D). The *HPT* gene shares *3′UTR* with the target *GFP* gene. So the 3′ transitive sRNAs from the target *T-DNA* could also target the *HPT* gene and cause 5′ transitivity into the *HPT* coding sequence (this was the only clear case of 5′ transitivity in our experimental system). This was also supported by the decrease in *HPT* mRNA

levels in AS and UT lines after the induction of silencing (Fig. S5I). The *NPTII* (kanamycin phosphotransferase) gene in the target *T-DNA* also shares the same *3′UTR* as the target *GFP* gene and the *HPT* gene, nevertheless, there were no secondary sRNAs spreading into the *NPTII* gene (Fig. S5A, C). However, like with the *HPT* gene, mRNA levels of the *NPTII* gene also decreased when the lines were exposed to estradiol (Fig. S5I). In the UT3 line, there were also high levels of sRNAs matching to the 3′ part of the *XVE* gene encoding the estradiol receptor (Fig. S5B). Most of these sRNAs appeared only after induction and they behaved the same as the *GFP* sRNAs. These sRNAs were probably related to the presence of the direct tandem repeat in this line, as described above (Fig. S4).

In the IR variant, the amounts of sRNAs that matched regions outside the *GFP* sRNA source sequence were extremely low (Fig. S5C). One of the regions with low levels of sRNAs was the linker between the inverted *GFP* sequences. When analyzing the sRNA data, we uncovered that the line IR1 has a rearrangement in the linker between *GFP* sequences, which is much larger than it should be (Fig. S5J). Sequencing of the linker showed that the sequence between the exact end of the first *GFP* in the repeat to the donor splice site was replaced with 839 bp from the central part of the *XVE* gene and 1514 bp from the backbone of pER8 plasmid, starting with the end of the left border region (Fig. S5J). Though unintended, the fact that the linker contained part of the sequence that was not transcribed from any other locus and part of a sequence that was transcribed within the *XVE* transcription unit, gave us an opportunity to analyze the sRNAs from this region in more detail (Fig. S5K, L). The sRNAs from the region homologous to the *XVE* reached higher levels (3-fold) than those in the surrounding regions. They also differed in the representation of sRNAs of different lengths (32% of 22 nt from region matching *XVE*, compared to 45% of those from the surroundings; Fig. S5L). Even though the levels of sRNAs from the linker reached only low levels, they were able to induce weak 3′ transitivity on the *XVE* transcript (Fig. S5C).

2.4.2. DNA methylation accompanying the PTGS

Presence of sRNAs is known to induce DNA methylation. To see whether our silencers differed in this aspect we performed bisulfite sequencing of two regions encompassing approximately 600 bp from the *35S* promoter and 600 bp from the *GFP* coding sequence (Figs. 6A, B; S6A, B) at three key time points in the experiment (0, 14 and 35 days). There was no DNA methylation during the whole experiment in the AS variant. In contrast, the IR was able to induce high levels of DNA methylation in all sequence contexts along the whole analyzed target *GFP*. Specifically, the methylation in the CHG context reached 96% and when looking at the CWG context alone the methylation was 100% (there were 14 CWG cytosines in that region). There was also a weak spill-over of methylated cytosines outside of the target *GFP* at the 5′ end, that reached about 40 nt upstream (we have no data from the 3′ end of the target *GFP*). After removal of estradiol, the methylation in the CG context remained at the same level, but the methylation in CHG and CHH context almost fully disappeared (Figs. 6A; S6B).

In the UT3 line we chose for detailed analyses, DNA was methylated in the CG context already at time 0, which may reflect some transient silencing that might have occurred in this particular line before the experiment began. After induction, there was a weak increase in CHG and CHH methylation at the 3′ end of the *GFP* sequence, but a decrease in CG methylation in the middle of the target *GFP* (the changes in both CG and CHH were statistically significant, $p < 0.05$).

Though our previous study showed that preexisting CG methylation did not influence de novo methylation [39], we included one additional line (UT2) to our analysis. This line was free of the initial CG methylation, and the estradiol induced methylation in all three sequence contexts. This methylation was limited to the 3′ end and remained almost unchanged even after estradiol removal (Fig. S6C).

To verify the results of the bisulfite sequencing in all three lines of each variant, we used McrBC assay that is based on qPCR quantification of DNA digested with an enzyme that specifically cuts methylated DNA. The results showed that methylation in all the AS lines was very weak or missing. In contrast, all three IR lines showed strong induction of methylation that remained high even after estradiol removal, likely in the form of CG methylation (Fig. 6C). The UT lines differed among themselves. As expected, the UT3 line with preexisting CG methylation showed consistently high methylation, while the UT2 line showed estradiol-induced methylation and the UT1 line behaved somehow in between.

We also used the McrBC assay to analyze the methylation at the

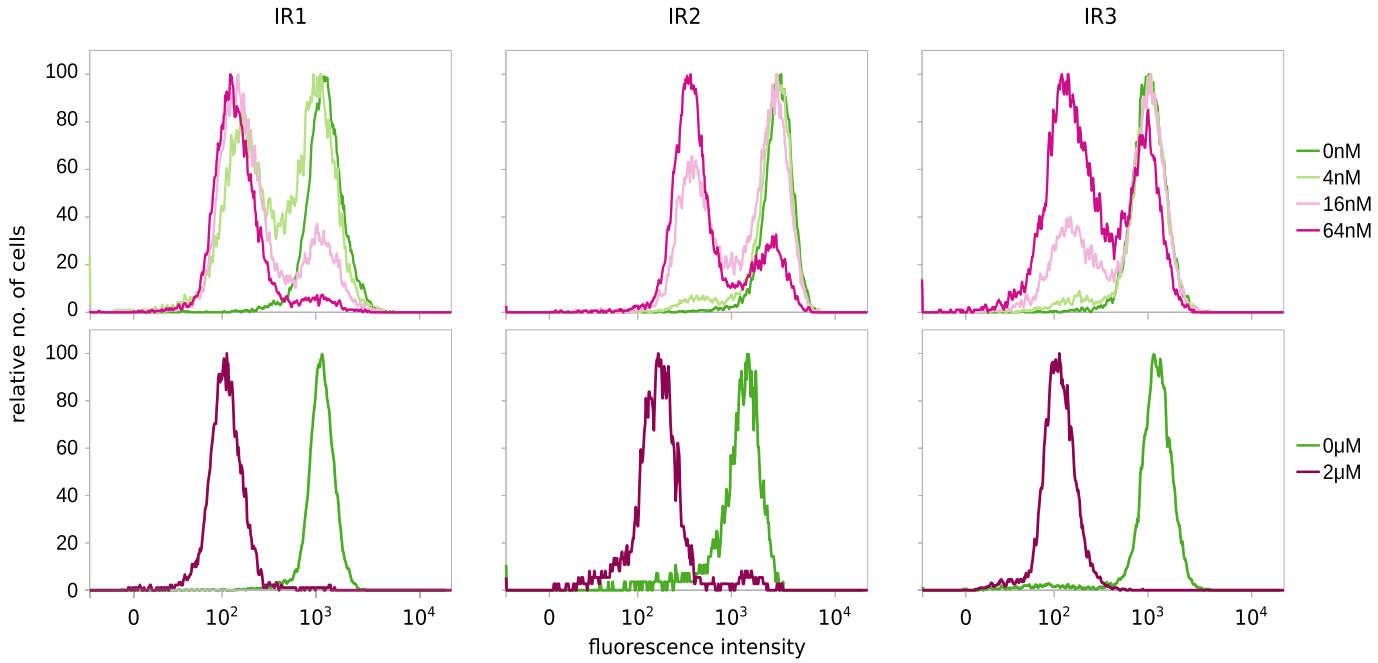


Fig. 4. Effects of estradiol concentration on the induction of silencing after 3 days.
Histograms of cell GFP fluorescence levels (flow cytometer data normalized to mode) showing binary silencing at different estradiol concentrations in three different IR lines. The second line of graphs shows data for fully induced lines (2 μM estradiol) from the experiment documented in Fig. 3.

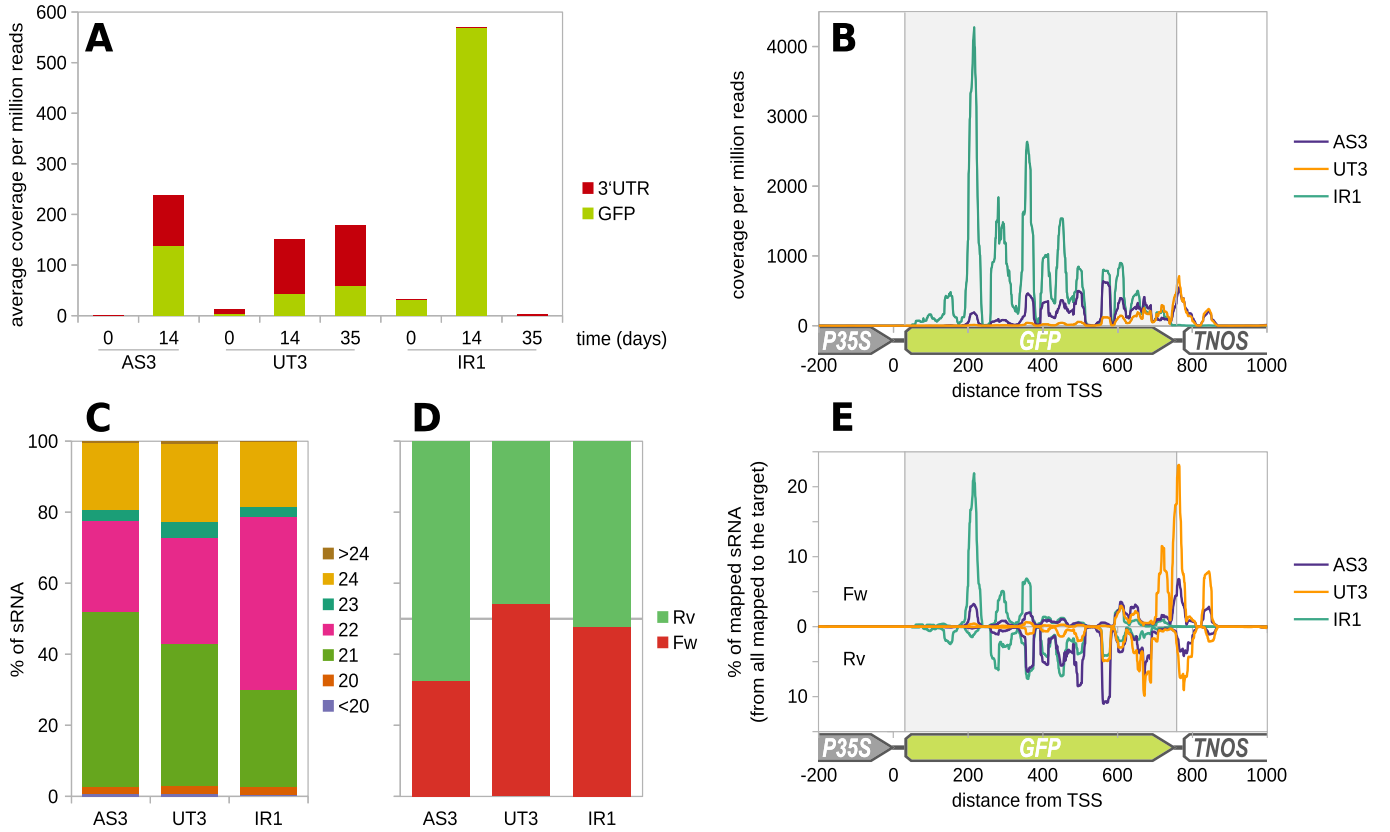


Fig. 5. Characterization of sRNAs matching with the target *T-DNA*.
Small RNA levels were assessed in suspension cultures treated with β–estradiol for 14 days and recovered for an additional 21 days (see Fig. 3 for the time course of silencing).
(A) Average coverage of the target *GFP* and its 3′ untranslated region (*3′UTR*) with sRNAs.
(B) Distribution of sRNAs of all size categories along the target *GFP* and its 3′ untranslated region at day 14 (more detailed analyses are presented in Fig. S5G).
(B) Distribution of sRNAs of all size categories along the target *GFP* and its 3′ untranslated region at day 14 (more detailed analyses are presented in Fig. S5G).
(C) Length distribution of sRNAs matching to the target *T-DNA* at day 14.
(D) Strand distribution of sRNAs matching to the target *T-DNA* at day 14 (Fw = forward, Rv = reverse).
(E) Distribution of sRNAs of all size categories along the target *GFP* and its 3′ untranslated region at day 14 based on their strand orientation (Fw are above X axis, Rv are below X axis). Each line is on a different scale for easier comparison of the position of individual peaks (more detailed analyses are presented in Fig. S5E).
(E) Distribution of sRNAs of all size categories along the target *GFP* and its 3′ untranslated region at day 14 based on their strand orientation (Fw are above X axis, Rv are below X axis). Each line is on a different scale for easier comparison of the position of individual peaks (more detailed analyses are presented in Fig. S5E).

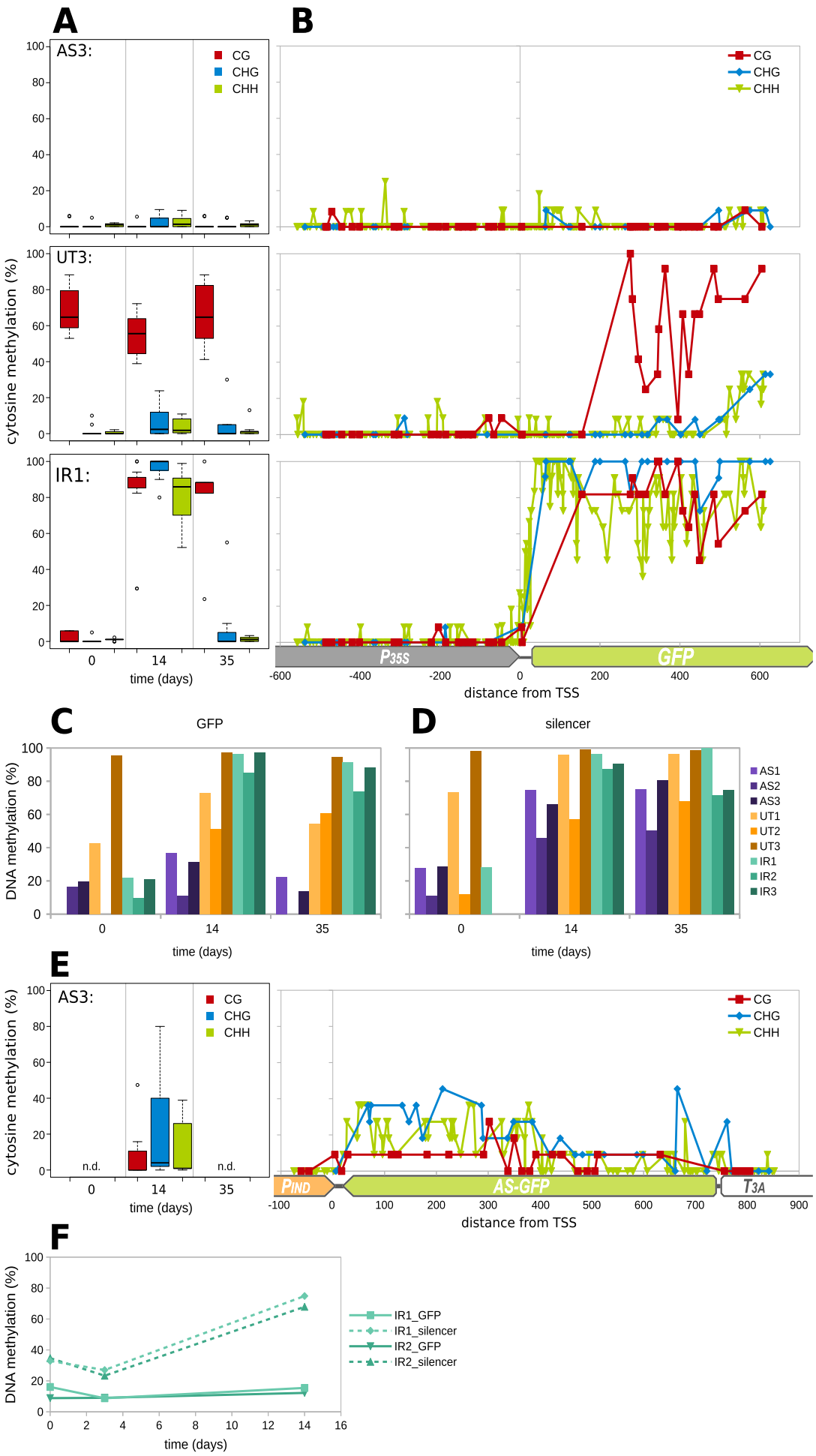
silencer locus. For the UT and IR variants, the behavior was highly similar to the situation at the target locus, but the methylation differed in the case of the AS variant (Fig. 6D). Unlike the target region, the AS expression was able to induce some DNA methylation at the silencer locus. To verify this result, we did bisulfite sequencing of the AS silencer on the 14th day after induction from the same sample that we used for the target *GFP* analysis (Fig. 6E). Indeed, we did detect mid-levels of DNA methylation, especially in the CHG context, but also weak methylation in the CHH and CG context along the whole antisense *GFP* sequence. Using the McrBC assay we could also see that detectable changes in methylation in the IR1 line appeared after two days at the target *GFP*, but as early as after one day at the silencer locus (Fig. S6D). Moreover, it turned out that the weakly induced IR was not able to induce DNA methylation at the target locus, although it was still able to induce weak methylation at the silencer locus (Fig. 6F).

3. Discussion

In this study, we closely analyzed the effects of different silencers (sources of sRNAs) on PTGS, its strength, efficiency, timing, dynamics, reversibility, and accompanying processes like DNA methylation, secondary sRNA production and transitivity. All silencers were based on identical sequence, full-length *GFP*, and the silencing target was constitutively and stably expressed *35S::GFP* in the identical locus, allowing direct comparisons of the silencing features. Our results showed consistent behavior in all selected IR and AS lines indicating that their character was indeed connected with different sRNA sources. In case of UT silencer, the best responding lines were more heterogeneous possibly due to involvement of other RNAi pathways that might be related to complex T-DNA insertions in these lines.

3.1. Silencing strength, efficiency and dynamics differed among silencers

Our data confirmed previous observations that IR is the most effective inducer of silencing (Table 1; Fig. S1) most likely thanks to the highest efficacy of dsRNA formation leading to very high levels of sRNAs. IR silencer caused the strongest reduction of target gene expression (as low as 2% of the original transcript level; Fig. S2A, B; [13,33–35]). The efficiency that we observed for IR was lower than usually obtained, which likely resulted from the fact that about 70% of IR lines had silenced GFP expression spontaneously (before induction with estradiol) due to read-through transcription from the adjacent genomic region into the *T-DNAs* [38]. These lines, which were not included in our analysis, would have been probably classified as successfully silenced when assessing the efficiency with a constitutively expressed silencer. Silencing induced by UT [32,43,44] was less frequent in our experimental system and did not significantly differ from classical cosuppression (induced by terminated sense *GFP* transcripts;



(caption on next page)

Fig. 6. DNA methylation of the GFP target and silencer during PTGS.
DNA methylation levels were assessed in suspension cultures treated with β–estradiol for 14 days and recovered for an additional 21 days (see Fig. 3 for the time course of silencing).
(A) Boxplots showing average DNA methylation levels in CG, CHG and CHH contexts in the analyzed region of the target *GFP* gene. Analysis was done by bisulfite modification in one line representing each silencer variant (lines AS3, UT3 and IR1), 8 to 12 DNA clones were sequenced for each data point.
(B) Distribution of methylated cytosines in the three sequence contexts along the target region and in the 5′ adjacent region after 14-day β–estradiol treatment.
(C) Methylation level at the target *GFP* gene and (D) at the *GFP* region of the silencer *T-DNA* assessed by McrBC assay (primers for qPCR are listed in Table S1).
(C) Methylation level at the target *GFP* gene and (D) at the *GFP* region of the silencer *T-DNA* assessed by McrBC assay (primers for qPCR are listed in Table S1).
(E) Boxplots showing levels of DNA methylation in CG, CHG and CHH contexts for the silencer *GFP* region in AS3 line at 14th day of induction, and distribution of methylated cytosines along the silencer *GFP* region (including 3′ and 5′ adjacent regions) for the same data (data from 11 sequenced DNA clones).
(F) Target GFP and silencer methylation over time as assessed by McrBC assay for lines IR1 and IR2 at 16 nM concentration of estradiol.

Table 1; Fig. S1) challenging the impact of the missing terminator for silencing induction.

With data from the inducible system, we could extend the comparisons beyond just silencing strength and efficiency. The silencing with IR was faster when compared to UT and AS (with the first reaction within 6 h; Figs. 3; S2) similar to the 3 to 24 h that was previously observed using inducible silencing with IR in *Arabidopsis* and tobacco [45–47].

When testing the impact of the strength of silencing induction, we found that at the cell level the silencing was not continuous in the IR lines, but the cells were either silenced or not (Fig. 4). This might hint to some mechanisms that perhaps serve to prevent sporadically produced sRNAs from inducing unwanted silencing. Only when sRNA level reaches a certain threshold then they will have the potential to induce silencing [48].

In the UT lines, one would expect that silencing is strongly dependent on the intensity of induction with some threshold level because UT silencing is believed to be dependent on production of aberrant unterminated transcripts. Such transcripts could enter the RNAi pathway only after they saturate the RNA degradation pathways [32], so the UT construct should perform much worse or completely fail at low induction levels. Surprisingly, it performed much better than AS and, given enough time, it almost matched the IR (Fig. 2). This low dependence of UT silencing on the inducer level and the same performance of UT and GF lines in the population callus screen suggest that the mechanism of silencing in the UT lines might be different from that we expected. Recently, Parent et al., 2015 [49] documented that S-PTGS in *Arabidopsis* L1 line (line with spontaneous silencing of GUS transgene) was connected with the formation of aberrant read-through antisense RNAs from the oppositely oriented *NPTII* transgene. Though it might be considered as a specific situation, a similar mechanism seems to be important in the regulation of a large number of *Arabidopsis* genes [49]. Other studies also linked cosuppression to the read-through transcription [50] and some of the important works on S-PTGS were done on plants with convergently oriented transgenes including the work of Luo and Chen [32]. Our results show that in lines carrying the UT construct, aberrant untranslatable RNAs arose more frequently (by at least 30%) compared to the control terminated GF lines (Fig. S1C), but the frequency of silencing did not differ between the UT and GF variants (9% in both). It suggests that aberrant unterminated RNAs in our UT lines were probably not sufficient for the induction of efficient silencing. More likely, the transcripts generated from the inducible promoter required other features for silencing induction, perhaps similar to the situation described in the *Arabidopsis* L1 line, where silencing was connected with the antisense read-through transcription ([49]; Fig. S4G). This read-through transcription might result from the complex arrangement of *T-DNA* insertions or from the genomic context as discussed previously [38,49]. As such, it might be much more difficult or even impossible to induce such silencing without more complex T-DNA structure.

3.2. Silencing with UT could not be fully recovered

IR silencing in our experiments showed practically full transcriptional reactivation after the removal of the inducer consistently with previous studies [45–47]. The same was observed for the AS variant, but in the UT2 and UT3 lines GFP expression could not be fully recovered after estradiol removal (Figs. 3; S2). The silencing maintenance was unlikely caused by promoter leakiness because in all analyzed lines the silencer transcript levels at day 35 were lower than the background levels detected at day 0 (Fig. S2C). We also did not observe any DNA methylation spreading into the *35S* promoter (Fig. 6B), so it was unlikely that the persisting silencing was caused by TGS. PTGS maintenance without the inducer was already observed after viral infection [51] or transient agrobacterium transformation [52]. Also, classical S-PTGS that has features in common with UT silencing can be maintained when spontaneously initiated [53]. However, specific features of this silencing were, to our knowledge, never described in detail. Here we show that the maintenance of silencing in the UT3 line was accompanied by the continual presence of sRNAs that even slightly increased in abundance after the inducer removal (Fig. 5A). Stable sRNA levels after inducer removal can also be inferred for the UT2 line from persisting CHH methylation. Interestingly, there were no evident differences between the profile and composition of sRNA for the UT and AS lines that could easily explain this difference in their behavior (Figs. 5B, E; S5B, C, E). The *3′UTR* sRNAs for the AS and UT lines were virtually identical and their ability to induce production of transitive sRNAs from the HPT gene was the same. There were some differences between the sRNAs matching to the *GFP* region, however, the UT sRNAs in this region seemed to be mostly a subgroup of sRNAs present in the AS line. As such, the reason for the maintenance of the UT silencing likely lies somewhere else, like in the source or in the site of sRNA production. Nevertheless, the inducible UT construct could provide a tool for studying the mechanisms involved in the maintenance of PTGS.

3.3. UT and AS silencing was accompanied with strong sRNA transitivity

Sequencing of sRNAs identified large amounts of secondary sRNAs in the AS and UT line. In the UT line, the transitive sRNAs against *3′UTR* made up 41% of the sRNAs targeting the *GFP* transcript (Fig. 5A, B). Moreover, after the removal of the inducer, there was no decrease in their amount and no change in their profile (Figs. 5A; S5E), suggesting that the primary sRNAs made up only a very small fraction of the sRNAs targeting the *GFP* transcript. Similar amounts of transitive sRNAs were also present in the AS line (Fig. 5A, B). Also based on a single point mutation in the AS silencer, secondary sRNAs made up at least 60% of sRNAs overlapping that position (Table S3B). In both the AS and UT line, the *3′UTR* secondary sRNAs were also able to induce production of secondary 5′ transitive sRNAs in trans from the *HPT* gene, but interestingly not from the *NPTII* gene with an identical terminator (Fig. S5C). This could have resulted from different sequences present upstream of the T_{NOS} part of the *3′UTR* sequence, or it could have resulted from different distances between the sRNA recognition site and the end of the coding sequence between both genes (29 nt for *HPT* and 387 nt for *NPTII*; Fig. 1). It has been previously shown that translation and production of secondary sRNAs are closely connected [54,55]. The *HPT* gene is also the only clear example of 5′ transitivity in our system. This might be because in our system sRNAs mostly targeted coding sequences, where the 5′ degradation products could be efficiently degraded by the nonstop decay (NSD) pathway [56], whereas in the case

of *HPT* gene, sRNAs targeted the untranslated region.

3.4. IR silencing was not accompanied with transitivity despite high level of 22 nt sRNAs

While in the AS and UT lines there was strong 3′ transitivity on the *GFP* transcript and the majority of sRNAs seemed to be secondary (Fig. 5A, B), in the IR line, there was no evidence of the production of significant amounts of transitive sRNAs. Only rare sRNAs from the linker between the arms of the inverted repeat were able to induce low levels of transitive sRNAs on the *XVE* transcript (Fig. S5K). Missing 3′ transitivity is in contradiction with what has been shown previously, that inverted repeats were able to induce strong 3′ transitivity, especially when targeting transgenes [57–60]. Also, unlike the situation in our system, most inverted repeat sRNAs are dominated by 21 nt fraction of sRNAs, because DCL4 is the lead dicer in processing long dsRNA [3,60–62]. This is a characteristic also common in most endogenous IRs in *Arabidopsis*, except for IR71, which is dominated by 22 nt sRNA [63]. Dalakouras et al. observed silencing of GFP induced by an inverted repeat, where the majority of sRNAs were 22 nt long and there were only low levels of transitive sRNA [64], resembling our case (Fig. 5B, C). They speculated that this was caused by the high expression of DCL2 in *N. benthamiana*. The same might also apply to the *Nicotiana* BY-2 cells, but not generally, because when we expressed *CaMV 35S* promoter hairpin in BY-2 cells, the sRNA size classes were similar to other IRs, with the dominant fraction being 21 nt long [39].

3.5. AS silencing showed strong sRNA strand asymmetry

Data from sRNA sequencing showed an extremely shifted ratio of forward and reverse strand sRNAs in the AS line (Fig. 5D). The preference for the sRNAs to originate from the forward or the reverse strand is locally affected by the sequence composition, as can be seen from the distribution of peaks of sRNA coverage along the sequence (Fig. 5E). On a larger scale, the strand ratio is almost even for the whole *GFP* transcript in the UT and IR lines. The shift in the ratio favoring antisense strands observed in AS line was mostly caused by decreased sRNA production from two regions with a strong bias for forward strand sRNAs at the 5′ and 3′ end of the *GFP* coding sequence compared to the IR and UT lines, respectively (Figs. 5E; S5G). The stronger production of sRNAs at the 5′ end in the IR line was probably caused by the way the hairpin RNA is processed. It has been previously shown that more sRNAs are produced from regions closer to the loop [59], which is the 5′ end of the *GFP* sequence in our case. However, at the 3′ end the AS line seemed to be able to produce sRNAs well enough, but the production was lower only in one specific location, with the highest bias towards the forward strand, affecting all types of sRNAs (Fig. S5G). To be able to explain what caused this difference, it would be necessary to know which sRNAs are primary and which are secondary. Though the mechanism responsible for the strand bias remains unclear, it might be a general feature of processing dsRNAs that originate from intermolecular pairing, as it is common among nat-siRNAs to have a highly uneven sRNA strand ratio [65].

3.6. Loci with the same sequence can show different sensitivity to DNA methylation

The AS variant also showed unexpected differences in its ability to methylate corresponding DNA sequences. All the AS lines could clearly methylate the sequence of the silencer, but they were unable to induce significant changes in methylation at the target *GFP* (Fig. 6C, D). To explain this difference between methylation of the target and the silencer locus in AS lines, one might look for a connection with the strand asymmetry in sRNA production. But it is unlikely the cause since weakly induced IR lines behaved in the same way (Fig. 6F). So, we prefer an alternative explanation. The GFP from the target locus has been stably expressed for several years [37], whereas the expression from the silencer locus has been switched on right at the start of the experiment. It is reasonable to expect that these loci differ in their epigenetic state. The locus with more chromatin activation marks could be more resistant to deposition of repressive marks, as was reported in an animal system [66]. Alternatively, the stronger ongoing transcription could more effectively erase newly deposited repressive chromatin marks [67]. It has been also previously observed that transgenes inserted into active genes are more resistant to DNA methylation [68] or that endogenous plant genes are more resistant to DNA methylation than transgenes [69].

Interestingly no such difference was observed in the UT lines, where the production/level of sRNAs was even lower than it was for the AS line. The overall amount of 24 nt sRNAs, which are primarily responsible for DNA methylation, was in the UT3 line six-times lower compared to the AS3 line over the region where we analyzed the methylation (the relative proportion of 24 nt sRNAs was similar among the variants; Fig. 5C). This suggests that there might be some qualitative difference between AS and UT sRNAs, which is not obvious from their length or sequence. The likely explanation is that in the UT lines, most of the sRNAs are produced in the nucleus (at least the primary sRNAs). The aberrant transcripts lack a polyA tail, so they are not exported out of the nucleus and must be processed there [41]. In the AS lines, on the other hand, both sense and antisense transcripts could be exported to the cytoplasm, so it is possible that most of the sRNAs were produced in the cytoplasm. Such sRNAs would have a harder time accessing the DNA and it would be more difficult for them to induce DNA methylation. On the other hand, most DCLs seem to be localized in the nucleus [2,70–72] and AGO4 loads sRNAs in the cytoplasm [73], so it remains an open question as to what effect the localization of dsRNA and sRNA production have on their activity. The difference between the ability of AS and UT variants to induce DNA methylation was not caused by the preexisting DNA methylation in the UT lines (Figs. 6A; S6A). The UT2 line was not methylated before induction and yet it was easily methylated after induction (Figs. 6C; S6C). Also, we have previously shown that preexisting CG methylation does not enhance methylation by RdDM at all [39].

4. Conclusions

In this work, we compared silencing induced by three different variants of silencing inducers which used different ways to form dsRNAs. Though we supposed that the mechanisms of silencing by the three silencers were well-established, data obtained in our system indicated that in case of the UT silencing, the production of aberrant unterminated RNAs was probably not sufficient to induce silencing, but other features connected with the T-DNA insertions were also required. These might be similar to the S-PTGS that was reported to be induced by read-through antisense transcripts [49].

The three silencers we tested varied considerably regarding silencing dynamics, reversibility, transitivity, and accompanying DNA methylation. The induction of silencing was fastest and strongest with the IR silencer. The AS silencing was most dependent on the induction level. The silencing was maintained after inducer removal specifically in some UT lines. Small RNAs reached the highest levels in the IR variant and lowest in the UT one. Both the AS and UT lines were able to induce strong production of transitive sRNAs, unlike the IR. Surprisingly, in the AS variant, sRNAs showed strongly deviated sRNA-strand ratio shifted in favor of antisense orientation, but the mechanism of it remains unclear.

Our results on DNA methylation accompanying PTGS suggested that there are several factors influencing this phenomenon; i) the strength of induction likely connected with the amount of sRNAs, because IR-induced methylation was strongly dependent on the inducer level, ii) the origin of sRNAs, because sRNAs were present in several times higher levels in the AS variant compared to the UT variant, but the GFP gene was methylated only in the UT variant, iii) the state of the locus targeted by sRNAs, because the silencer GFP was methylated, but the target GFP was not methylated in the AS lines and in the low inducted IR lines

despite the sequence identity of these loci. We assume that this sensitivity was affected by the epigenetic state of the locus (e.g. presence of certain chromatin marks), but further research is needed.

Even though we targeted the same sequence at the same locus with the silencing inducers, all of which were based on the same whole *GFP* coding sequence, the silencing responses were surprisingly variable. These differences in behaviors were likely connected with the varying mode of dsRNA precursor formation. Thus, our results raise several questions about specific aspects of RNAi that still await answers.

5. Materials and methods

5.1. Plant material

Tobacco cell line BY-2 (*Nicotiana tabacum* L. cv. Bright Yellow 2; [36]) was cultivated on modified MS medium with 3% sucrose [74] in darkness at 26 °C. The calli were grown on solidified medium (0.8 w/v agar; 6 cm diameter plates) and subcultured monthly by small pieces (3–5 mm diameter). The suspension cell cultures were cultured in 30 ml of medium in 100 ml Erlenmeyer flasks on an IKA KS501 orbital shaker at 110 rpm (IKA Labortechnik, Staufen, Germany; orbital diameter 30 mm).

The suspension cell cultures were kept in the exponential phase for all experiments by subcultivation every 3–4 days with 2–3 ml of the cell culture (i.e. about 100 mg of fresh cell biomass).

Transformations of BY-2 suspension cells were carried out as described previously [38] using *Agrobacterium tumefaciens* strain C58C1 carrying helper plasmid *pGV2260* and an appropriate binary vector (Fig. 1; for details on construct preparation see [38]). After cocultivation with agrobacterium, the cells were plated on solidified medium containing 25 μg/ml hygromycin and 100 μg/ml cefotaxim and cultured for 3 weeks. Using this procedure, individual transformed cells formed isolated macroscopic cell clusters (commonly called calli) that could be mostly regarded as genetically homogeneous clones [37].

The XVE inducible system [40] was activated by cultivating the calli or cell suspension cultures on media with an addition of 2 μM β-estradiol (Sigma-Aldrich Cat. No. E2758) unless otherwise specified. The β-estradiol was stored as 20 mM solution in DMSO, therefore a corresponding amount of DMSO was added to the cultivation medium of controls. To remove the β-estradiol from the cell culture (on day 14 of the experiments), 3 ml of the cell culture were washed (diluted and filtered) three times with 50 ml of 3% (w/v) sucrose and then resuspended in a fresh MS medium (30 ml).

5.2. Fluorescence analysis

For population analysis of BY-2 calli, seven calli per plate were grown for 14 days. For each line, one callus was cultivated on medium with estradiol (induced) and one was cultivated on medium without estradiol (mock-induced). Fluorescence was measured every 2–3 days. Each plate was separately photodocumented using G:BOX (SynGene, Cambridge, UK) with a blue excitation light (LED diodes with maximum at $\lambda = 465$ nm) and a green emission filter (FILT525/GX; 510–540 nm). The images were processed using software NIS-Elements 3.10 (Laboratory Imaging, Prague, CZ). The average light intensity was measured for all the pixels from each callus. Lines that strongly differed in callus fluorescence between the induced and mock-induced variants already at time 0, i.e. lines whose fluorescence did not match the formula: $|\log(|\text{ind./mock-ind.}|)| < |\text{average}(\log(\text{ind./mock-ind.}))| + 2 \times \text{stdev}$ due to non-homogeneous character of the source callus, were excluded from further analyses. Background fluorescence (measured from WT calli) was subtracted from the fluorescence of each callus. After that, the fluorescence of each line was normalized to its mock-induced variant: (induced/mock-induced) × 100. Statistical analyses were done by Wilcoxon signed-rank test and Wilcoxon rank-sum test in R 3.4.4.

Fluorescence from cell suspension cultures was measured by flow cytometry. First, protoplasts were prepared by incubating 50–100 mg of cells in 1.5 ml of protoplast enzyme solution: 0.5% (w/v) cellulase R-10 (Duchefa), 0.1% (w/v) pectolyase Y-23 (Duchefa) in 0.45 M mannitol (Sigma-Aldrich) for 3 h at 26 °C with gentle shaking. Protoplasts were centrifuged (200 ×*g* for 5 min) and the pellet was gently resuspended in MS medium with 0.4 sucrose. The samples were then measured on LSR II flow cytometer (BD biosciences). On average 14,000 cells were measured per sample. The data were processed by Flowing Software (http://flowingsoftware.btk.fi/) with filtering live protoplasts as described in [75].

5.3. Transcript analyses

RNA was extracted from 100 mg of cells using the phenol-chloroform extraction method [76]. The integrity of the RNA for each sample was checked by gel electrophoresis. For cDNA preparation, 1 μg of the total RNA was treated with DNase I and half of the reaction mixture was then used as a template for RevertAid Reverse Transcriptase (Fermentas, Thermo Fisher Scientific). The other half was subjected to the same treatment without the addition of the reverse transcriptase to check for DNA contamination. A mixture of specific primers was used for the reverse transcription (see Table S1). The final reaction was diluted to 60 μl (template).

The qPCR was done using LightCycler 480 (Roche) and iQ™ SYBR® Green Supermix (BioRad, Hercules, CA, USA) in triplicates of 10 μl reactions with 1 μl of template. We adhered to the qPCR guidelines throughout the experiments [77]. The specificity of the qPCR was verified by melting curve analysis (using the LightCycler 480 software) and also by checking randomly selected samples using gel electrophoresis. The resulting data were processed by LinRegPCR 2015.3 [78], allowing for the correction of the amplicon amplification efficiency. Calculated transcript concentrations were normalized to the *NtEF1α* transcript level, so all the presented values show the relative level of a given transcript to the level of *NtEF1α*. Primers for qPCR are listed in Table S1. Some primer sequences were taken from previous studies [38,79].

5.4. DNA methylation analyses

5.4.1. Bisulfite conversion

The bisulfite conversion was done as described previously [80]. In brief: DNA was extracted using DNeasy Plant Mini Kit (Qiagen) from 100 mg of cells and digested with *Eco*RI. The bisulfite conversion of 600 ng of purified DNA was performed according to the EpiTect Bisulfite Kit (Qiagen). Primers designed to anneal to bisulfite modified DNA (Table S1) were used via PCR to amplify the regions of interest. The PCR products were cloned into *pDrive* vector (QIAGEN PCR Cloning Kit) and 8–12 clones per sample were sequenced. The resulting sequences were then processed using spreadsheet formulas in LibreOffice 6.1 and the statistical analysis was done using the Wilcoxon rank-sum test in R 3.4.4.

5.4.2. The McrBC assay

The level of methylation was further confirmed by qPCR after cleavage of genomic DNA with McrBC endonuclease (New England Biolabs) specific for methylated DNA (modified from [81]). In brief: 100 ng of DNA was first fragmented with *Ase*I that does not cut in the regions of interest and then the reaction was split in half and supplemented with 10 units of McrBC enzyme or an equivalent amount of 50% (v/v) glycerol. The DNA was digested for 6 h at 37 °C, and then the enzyme was inactivated (20 min, 65 °C). One nanogram was used for qPCR performed as described above with primers listed in Table S1.

5.5. Small RNA analyses

Total RNA was extracted from 100 mg (fresh weight) of cells using RNeasy Plant Mini Kit (Qiagen), quality assessed and quantified. A

fraction of sRNAs ranging in size from 18 to 45 nt were excised and recovered from 15% urea-polyacrylamide gels. Extracted sRNAs were ligated with 5′ and 3′ RNA adapters with T4 RNA ligase. The adapter-ligated small RNAs were subsequently transcribed into cDNA by Super-Script II Reverse Transcriptase (Invitrogen) and amplified using adaptor-specific primers. The amplified cDNA products were size-purified and circularised (ssDNA circles). This sRNA library was sequenced using the combinatorial probe-anchor synthesis (cPAS) based BGISEQ-500 sequencer (BGI, Shenzhen, China), which was previously shown to provide highly reproducible results comparable with other NGS platforms [82]. Obtained data were filtered using Geneious 9.1.8 (Biomatters, Ltd.) for reads perfectly matching all the *T-DNAs*. When filtering reads for the AS variant, a letter W (A or T) was placed at the position of the mutation in the sequence (404. nucleotide from the start of *GFP* cds, see Results and Fig. S5H). Data were sorted (based on sequence length, strand orientation and identity of 5′ nucleotide) and sRNA coverage maps were generated using tools available at usegalaxy.eu [83] while allowing HISAT2 [84] to map each read to every matching position. The maps of coverage were further processed using LibreOffice 6.1 and normalized to the overall amount of sRNAs from the given sample. For the purpose of data presentation, most graphs use average coverage calculated from these maps to allow for comparison of sequence elements of different length. Transcription units (transcription start sites and polyadenylation sites) were defined according to literature data for P_{35S} (including P_{IND} that contains the minimal P_{35S}), P_{NOS} and T_{NOS} [85,86], and for T_{3A} PASPA software (http://bmi.xmu.edu.cn/paspa/index.html) was used to predict the polyadenylation site [87].

5.6. Southern blot analysis

The Southern blot hybridization was done as previously described [15] with few modifications. Briefly, the DNA was isolated from 150 mg (fresh weight) of BY-2 calli. Twenty micrograms of genomic DNA per sample was separately digested by enzymes *Kpn*I, *Nsi*I and *Ase*I (New England Biolabs). The digoxigenin-labelled probes were prepared by PCR with DIG-dUTP (Roche) according to manufacturer's instructions. Parts of *HPT* and *GFP* genes were amplified with primers listed in Table S1. Hybridization and immunodetection with a chemiluminescent substrate CDP-Star (Tropix) was done according to DIG Application Manual (Roche).

The Southern hybridization results were interpreted as follows: tandem *T-DNA* inserted as a direct repeat should give the same 5.1 kbp fragment with all three restriction enzymes and both probes, plus one fragment of unknown size; head-to-head inverted repeat should give 10.0 kbp fragment when digested with *Kpn*I, two fragments of unknown size with *GFP* probe or 7.4 fragment with *HPT* probe for *Nsi*I and two fragments of unknown size when digested with *Ase*I; tail-to-tail inverted repeat should give two fragments of unknown size with *Kpn*I, 2.8 kbp fragment with *GFP* probe or two fragments of unknown size with *HPT* probe for *Nsi*I and 6.2 kbp fragment when digested with *Ase*I.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2020.194647.

Abbreviations

AS	antisense *GFP*
D	DMSO/control
dsRNA	double stranded RNA
E	β-estradiol/induction
GF	control *GFP*
GFP	green fluorescent protein
IR	inverted repeat *GFP*
PTGS	posttranscriptional gene silencing
RNAi	RNA interference
S-PTGS	sense-transgene induced posttranscriptional gene silencing
sRNA	small RNA
TGS	transcriptional gene silencing
UT	unterminated *GFP*

Data statement

All data are accessible within the manuscript and its supporting materials. The sRNA-seq datasets generated and/or analyzed during the current study are available in the European Nucleotide Archive, PRJEB33605 (https://www.ebi.ac.uk/ena/data/view/PRJEB33605).

CRediT authorship contribution statement

VC conducted most of the experimental work and participated in all experiments. DT participated in the flow cytometry, RT-qPCR and DNA methylation analyses and also in writing the manuscript. AP participated in methylation analyses. VC together with LF designed the study, interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was supported by the Ministry of Education, Youth and Sports of Czech Republic (project IDs: LO1417 and LM2015042).

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

We thank Lena Hunt for language corrections.

References

[1] N.G. Bologna, O. Voinnet, The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis, Annu. Rev. Plant Biol. 65 (2014) 473–503, https://doi.org/10.1146/annurev-arplant-050213-035728.

[2] Z. Xie, L.K. Johansen, A.M. Gustafson, K.D. Kasschau, A.D. Lellis, D. Zilberman, S. E. Jacobsen, J.C. Carrington, Genetic and functional diversification of small RNA pathways in plants, PLoS Biol. 2 (2004), e104, https://doi.org/10.1371/journal.pbio.0020104.

[3] A. Deleris, J. Gallego-Bartolome, J. Bao, K.D. Kasschau, J.C. Carrington, O. Voinnet, Hierarchical action and inhibition of plant dicer-like proteins in antiviral defense, Science. 313 (2006) 68–71, https://doi.org/10.1126/science.1128214.

[4] I.R. Henderson, X. Zhang, C. Lu, L. Johnson, B.C. Meyers, P.J. Green, S.E. Jacobsen, Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning, Nat. Genet. 38 (2006) 721–725, https://doi.org/10.1038/ng1804.

[5] K.D. Kasschau, N. Fahlgren, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, J.C. Carrington, Genome-wide profiling and analysis of Arabidopsis siRNAs, PLoS Biol. 5 (2007), e57, https://doi.org/10.1371/journal.pbio.0050057.

[6] S. Mi, T. Cai, Y. Hu, Y. Chen, E. Hodges, F. Ni, L. Wu, S. Li, H. Zhou, C. Long, S. Chen, G.J. Hannon, Y. Qi, Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide, Cell. 133 (2008) 116–127, https://doi.org/10.1016/j.cell.2008.02.034.

[7] E.R. Havecker, L.M. Wallbridge, T.J. Hardcastle, M.S. Bush, K.A. Kelly, R.M. Dunn, F. Schwach, J.H. Doonan, D.C. Baulcombe, The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci, Plant Cell 22 (2010) 321–334, https://doi.org/10.1105/tpc.109.072199.

[8] Z. Zhang, X. Liu, X. Guo, X.-J. Wang, X. Zhang, Arabidopsis AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing, Nat. Plants. 2 (2016) nplants201649. doi:https://doi.org/10.1038/nplants.2016.49.

[9] A. Takeda, S. Iwasaki, T. Watanabe, M. Utsumi, Y. Watanabe, The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins, Plant Cell Physiol. 49 (2008) 493–500, https://doi.org/10.1093/pcp/pcn043.

[10] A.L. Eamens, N.A. Smith, S.J. Curtin, M.-B. Wang, P.M. Waterhouse, The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide

strand selection from microRNA duplexes, RNA. 15 (2009) 2219–2235, https://doi.org/10.1261/rna.1646909.
[11] L. Yang, G. Wu, R.S. Poethig, Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in Arabidopsis, Proc. Natl. Acad. Sci. 109 (2012) 315–320, https://doi.org/10.1073/pnas.1114673109.
[12] M.A. Matzke, R.A. Mosher, RNA-directed DNA methylation: an epigenetic pathway of increasing complexity, Nat. Rev. Genet. 15 (2014) 394–408, https://doi.org/10.1038/nrg3683.
[13] M.-B. Wang, P.M. Waterhouse, High-efficiency silencing of a β-glucuronidase gene in rice is correlated with repetitive transgene structure but is independent of DNA methylation, Plant Mol. Biol. 43 (2000) 67–82, https://doi.org/10.1023/A:1006490331303.
[14] H. Saze, T. Kakutani, Differentiation of epigenetic modifications between transposons and genes, Curr. Opin. Plant Biol. 14 (2011) 81–87, https://doi.org/10.1016/j.pbi.2010.08.017.
[15] E. Nocarova, Z. Opatrny, L. Fischer, Successive silencing of tandem reporter genes in potato (Solanum tuberosum) over 5 years of vegetative propagation, Ann. Bot. 106 (2010) 565–572, https://doi.org/10.1093/aob/mcq153.
[16] T. Pélissier, S. Thalmeir, D. Kempe, H.-L. Sänger, M. Wassenegger, Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation, Nucleic Acids Res. 27 (1999) 1625–1634, https://doi.org/10.1093/nar/27.7.1625.
[17] R. Lister, R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, J. R. Ecker, Highly integrated single-base resolution maps of the epigenome in Arabidopsis, Cell. 133 (2008) 523–536, https://doi.org/10.1016/j.cell.2008.03.029.
[18] H.R. Woo, T.A. Dittmer, E.J. Richards, Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis, PLoS Genet. 4 (2008), e1000156, https://doi.org/10.1371/journal.pgen.1000156.
[19] A.M. Lindroth, X. Cao, J.P. Jackson, D. Zilberman, C.M. McCallum, S. Henikoff, S. E. Jacobsen, Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation, Science. 292 (2001) 2077–2080, https://doi.org/10.1126/science.1059745.
[20] A. Zemach, M.Y. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer, D. Zilberman, The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin, Cell. 153 (2013) 193–205, https://doi.org/10.1016/j.cell.2013.02.033.
[21] C. Liu, F. Lu, X. Cui, X. Cao, Histone methylation in higher plants, Annu. Rev. Plant Biol. 61 (2010) 395–420, https://doi.org/10.1146/annurev.arplant.043008.091939.
[22] J. Du, L.M. Johnson, M. Groth, S. Feng, C.J. Hale, S. Li, A.A. Vashisht, J. Gallego-Bartolome, J.A. Wohlschlegel, D.J. Patel, S.E. Jacobsen, Mechanism of DNA methylation-directed histone methylation by KRYPTONITE, Mol. Cell 55 (2014) 495–504, https://doi.org/10.1016/j.molcel.2014.06.009.
[23] X. Li, C.J. Harris, Z. Zhong, W. Chen, R. Liu, B. Jia, Z. Wang, S. Li, S.E. Jacobsen, J. Du, Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation, Proc. Natl. Acad. Sci. 201809841 (2018), https://doi.org/10.1073/pnas.1809841115.
[24] J.A. Law, J. Du, C.J. Hale, S. Feng, K. Krajewski, A.M.S. Palanca, B.D. Strahl, D. J. Patel, S.E. Jacobsen, Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1, Nature. 498 (2013) 385–389, https://doi.org/10.1038/nature12178.
[25] L.M. Johnson, J. Du, C.J. Hale, S. Bischof, S. Feng, R.K. Chodavarapu, X. Zhong, G. Marson, M. Pellegrini, D.J. Segal, D.J. Patel, S.E. Jacobsen, SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation, Nature. 507 (2014) 124–128, https://doi.org/10.1038/nature12931.
[26] D. Cuerda-Gil, R.K. Slotkin, Non-canonical RNA-directed DNA methylation, Nat. Plants 2 (2016) 16163, https://doi.org/10.1038/nplants.2016.163.
[27] O. Borsani, J. Zhu, P.E. Verslues, R. Sunkar, J.-K. Zhu, Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis, Cell. 123 (2005) 1279–1291, https://doi.org/10.1016/j.cell.2005.11.035.
[28] V. Ambros, B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, T. Tuschl, A uniform system for microRNA annotation, RNA. 9 (2003) 277–279, https://doi.org/10.1261/rna.2183803.
[29] F. Vazquez, H. Vaucheret, R. Rajagopalan, C. Lepers, V. Gasciolli, A.C. Mallory, J.-L. Hilbert, D.P. Bartel, P. Crété, Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs, Mol. Cell 16 (2004) 69–79, https://doi.org/10.1016/j.molcel.2004.09.028.
[30] M. Ronemus, M.W. Vaughn, R.A. Martienssen, MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent RNA polymerase in Arabidopsis, Plant Cell 18 (2006) 1559–1574, https://doi.org/10.1105/tpc.106.042127.
[31] C. Napoli, C. Lemieux, R. Jorgensen, Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans, Plant Cell 2 (1990) 279–289, https://doi.org/10.1105/tpc.2.4.279.
[32] Z. Luo, Z. Chen, Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis, Plant Cell 19 (2007) 943–958, https://doi.org/10.1105/tpc.106.045724.
[33] C.-F. Chuang, E.M. Meyerowitz, Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana, Proc. Natl. Acad. Sci. 97 (2000) 4985–4990, https://doi.org/10.1073/pnas.060034297.
[34] S.V. Wesley, C.A. Helliwell, N.A. Smith, M. Wang, D.T. Rouse, Q. Liu, P.S. Gooding, S.P. Singh, D. Abbott, P.A. Stoutjesdijk, S.P. Robinson, A.P. Gleave, A.G. Green, P. M. Waterhouse, Construct design for efficient, effective and high-throughput gene silencing in plants, Plant J. 27 (2001) 581–590, https://doi.org/10.1046/j.1365-313X.2001.01105.x.
[35] H. Yan, R. Chretien, J. Ye, C.M. Rommens, New construct approaches for efficient gene silencing in plants, Plant Physiol. 141 (2006) 1508–1518, https://doi.org/10.1104/pp.106.082271.
[36] T. Nagata, Y. Nemoto, S. Hasezawa, Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants, Int. Rev. Cytol. 132 (1992) 1–30, https://doi.org/10.1016/S0074-7696(08)62452-3.
[37] E. Nocarova, L. Fischer, Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression, BMC Plant Biol. 9 (2009) 44, https://doi.org/10.1186/1471-2229-9-44.
[38] V. Čermák, L. Fischer, Pervasive read-through transcription of T-DNAs is frequent in tobacco BY-2 cells and can effectively induce silencing, BMC Plant Biol. 18 (2018) 252, https://doi.org/10.1186/s12870-018-1482-3.
[39] A. Přibylová, V. Čermák, D. Tyč, L. Fischer, Detailed insight into the dynamics of the initial phases of de novo RNA-directed DNA methylation in plant cells, Epigenetics Chromatin 12 (2019) 54, https://doi.org/10.1186/s13072-019-0299-0.
[40] J. Zuo, Q. Niu, N. Chua, An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants, Plant J. 24 (2000) 265–273, https://doi.org/10.1046/j.1365-313x.2000.00868.x.
[41] Y. Huang, G.C. Carmichael, Role of polyadenylation in nucleocytoplasmic transport of mRNA, Mol. Cell. Biol. 16 (1996) 1534–1542, https://doi.org/10.1128/mcb.16.4.1534.
[42] C.M. Hommelsheim, L. Frantzeskakis, M. Huang, B. Ülker, PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications, Sci. Rep. 4 (2014) 5052, https://doi.org/10.1038/srep05052.
[43] S.J. Nicholson, V. Srivastava, Transgene constructs lacking transcription termination signal induce efficient silencing of endogenous targets in Arabidopsis, Mol. Gen. Genomics. 282 (2009) 319–328, https://doi.org/10.1007/s00438-009-0467-1.
[44] M.A. Akbudak, S.J. Nicholson, V. Srivastava, Suppression of Arabidopsis genes by terminator-less transgene constructs, Plant Biotechnol. Rep. 7 (2013) 415–424, https://doi.org/10.1007/s11816-013-0278-z.
[45] S. Chen, D. Hofius, U. Sonnewald, F. Börnke, Temporal and spatial control of gene silencing in transgenic plants by inducible expression of double-stranded RNA, Plant J. 36 (2003) 731–740, https://doi.org/10.1046/j.1365-313X.2003.01914.x.
[46] C. Lo, N. Wang, E. Lam, Inducible double-stranded RNA expression activates reversible transcript turnover and stable translational suppression of a target gene in transgenic tobacco, FEBS Lett. 579 (2005) 1498–1502, https://doi.org/10.1016/j.febslet.2005.01.062.
[47] A. Wielopolska, H. Townley, I. Moore, P. Waterhouse, C. Helliwell, A high-throughput inducible RNAi vector for plants, Plant Biotechnol. J. 3 (2005) 583–590, https://doi.org/10.1111/j.1467-7652.2005.00149.x.
[48] P. Pisacane, M. Halic, Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi, Nat. Commun. 8 (2017) 15332, https://doi.org/10.1038/ncomms15332.
[49] J.-S. Parent, V. Jauvion, N. Bouché, C. Béclin, M. Hachet, M. Zytnicki, H. Vaucheret, Post-transcriptional gene silencing triggered by sense transgenes involves uncapped antisense RNA and differs from silencing intentionally triggered by antisense transgenes, Nucleic Acids Res. 43 (2015) 8464–8475, https://doi.org/10.1093/nar/gkv753.
[50] M. Kasai, M. Koseki, K. Goto, C. Masuta, S. Ishii, R. Hellens, A. Taneda, A. Kanazawa, Coincident sequence-specific RNA degradation of linked transgenes in the plant genome, Plant Mol. Biol. 78 (2012) 259–273, https://doi.org/10.1007/s11103-011-9863-0.
[51] M.T. Ruiz, O. Voinnet, D.C. Baulcombe, Initiation and maintenance of virus-induced gene silencing, Plant Cell 10 (1998) 937–946, https://doi.org/10.1105/tpc.10.6.937.
[52] O. Voinnet, P. Vain, S. Angell, D.C. Baulcombe, Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA, Cell. 95 (1998) 177–187, https://doi.org/10.1016/S0092-8674(00)81749-3.
[53] H. Vaucheret, C. Béclin, M. Fagard, Post-transcriptional gene silencing in plants, J. Cell Sci. 114 (2001) 3083–3091.
[54] C. Zhang, D.W.-K. Ng, J. Lu, Z.J. Chen, Roles of target site location and sequence complementarity in trans-acting siRNA formation in Arabidopsis, Plant J. 69 (2012) 217–226, https://doi.org/10.1111/j.1365-313X.2011.04783.x.
[55] M. Yoshikawa, T. Iki, H. Numa, K. Miyashita, T. Meshi, M. Ishikawa, A short open reading frame encompassing the microRNA173 target site plays a role in trans-acting small interfering RNA biogenesis, Plant Physiol. 171 (2016) 359–368, https://doi.org/10.1104/pp.16.00148.
[56] I. Szádeczky-Kardoss, T. Csorba, A. Auber, A. Schamberger, T. Nyikó, J. Taller, T. I. Orbán, J. Burgyán, D. Silhavy, The nonstop decay and the RNA silencing systems operate cooperatively in plants, Nucleic Acids Res. 46 (2018) 4632–4648, https://doi.org/10.1093/nar/gky279.
[57] A. Molnar, C.W. Melnyk, A. Bassett, T.J. Hardcastle, R. Dunn, D.C. Baulcombe, Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells, Science. 328 (2010) 872–875, https://doi.org/10.1126/science.1187959.
[58] E. Dadami, A. Dalakouras, M. Zwiebel, G. Krczal, M. Wassenegger, An endogene-resembling transgene is resistant to DNA methylation and systemic silencing, RNA Biol. 11 (2014) 934–941, https://doi.org/10.4161/rna.29623.

[59] T. Wroblewski, M. Matvienko, U. Piskurewicz, H. Xu, B. Martineau, J. Wong, M. Govindarajulu, A. Kozik, R.W. Michelmore, Distinctive profiles of small RNA couple inverted repeat-induced post-transcriptional gene silencing with endogenous RNA silencing pathways in Arabidopsis, RNA. 20 (2014) 1987–1999, https://doi.org/10.1261/rna.046532.114.
[60] C. Taochy, N.R. Gursanscky, J. Cao, S.J. Fletcher, U. Dressel, N. Mitter, M.R. Tucker, A.M. Koltunow, J.L. Bowman, H. Vaucheret, B.J. Carroll, A genetic screen for impaired systemic RNAi highlights the crucial role of Dicer-like 2, Plant Physiol. (2017) pp.01181.2017. doi:https://doi.org/10.1104/pp.17.01181.
[61] S. Mlotshwa, G.J. Pruss, A. Peragine, M.W. Endres, J. Li, X. Chen, R.S. Poethig, L. H. Bowman, V. Vance, DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis, PLoS One 3 (2008), e1755, https://doi.org/10.1371/journal.pone.0001755.
[62] J.-S. Parent, N. Bouteiller, T. Elmayan, H. Vaucheret, Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing, Plant J. 81 (2015) 223–232, https://doi.org/10.1111/tpj.12720.
[63] S. Polydore, M.J. Axtell, Analysis of RDR1/RDR2/RDR6-independent small RNAs in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of short interfering RNA loci, Plant J. 94 (2018) 1051–1063, https://doi.org/10.1111/tpj.13919.
[64] A. Dalakouras, A. Lauter, A. Bassler, G. Krczal, M. Wassenegger, Transient expression of intron-containing transgenes generates non-spliced aberrant pre-mRNAs that are processed into siRNAs, Planta. 249 (2019) 457–468, https://doi.org/10.1007/s00425-018-3015-6.
[65] T.J. Hardcastle, S.Y. Müller, D.C. Baulcombe, Towards annotating the plant epigenome: the Arabidopsis thaliana small RNA locus map, Sci. Rep. 8 (2018) 6338, https://doi.org/10.1038/s41598-018-24515-8.
[66] S.K.T. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S.-P. Lin, C.D. Allis, X. Cheng, T.H. Bestor, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, Nature. 448 (2007) 714–717, https://doi.org/10.1038/nature05987.
[67] A. Miura, M. Nakamura, S. Inagaki, A. Kobayashi, H. Saze, T. Kakutani, An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites, EMBO J. 28 (2009) 1078–1086, https://doi.org/10.1038/emboj.2009.59.
[68] U. Fischer, M. Kuhlmann, A. Pecinka, R. Schmidt, M.F. Mette, Local DNA features affect RNA-directed transcriptional gene silencing and DNA methylation, Plant J. 53 (2008) 1–10, https://doi.org/10.1111/j.1365-313X.2007.03311.x.
[69] L. Vermeersch, N. De Winne, J. Nolf, A. Bleys, A. Kovařík, A. Depicker, Transitive RNA silencing signals induce cytosine methylation of a transgenic but not an endogenous target, Plant J. 74 (2013) 867–879, https://doi.org/10.1111/tpj.12172.
[70] A. Hiraguri, R. Itoh, N. Kondo, Y. Nomura, D. Aizawa, Y. Murai, H. Koiwa, M. Seki, K. Shinozaki, T. Fukuhara, Specific interactions between Dicer-like proteins and HYL1/DRB- family dsRNA-binding proteins in Arabidopsis thaliana, Plant Mol. Biol. 57 (2005) 173–188, https://doi.org/10.1007/s11103-004-6853-5.
[71] P. Hoffer, S. Ivashuta, O. Pontes, A. Vitins, C. Pikaard, A. Mroczka, N. Wagner, T. Voelker, Posttranscriptional gene silencing in nuclei, Proc. Natl. Acad. Sci. 108 (2011) 409–414, https://doi.org/10.1073/pnas.1009805108.
[72] N. Pumplin, A. Sarazin, P.E. Jullien, N.G. Bologna, S. Oberlin, O. Voinnet, DNA methylation influences the expression of DICER-LIKE4 isoforms, which encode proteins of alternative localization and function, Plant Cell 28 (2016) 2786–2804, https://doi.org/10.1105/tpc.16.00554.
[73] R. Ye, W. Wang, T. Iki, C. Liu, Y. Wu, M. Ishikawa, X. Zhou, Y. Qi, Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes, Mol. Cell 46 (2012) 859–870, https://doi.org/10.1016/j.molcel.2012.04.013.
[74] T. Murashige, F. Skoog, A revised medium for rapid growth and bio assays with tobacco tissue cultures, Physiol. Plant. 15 (1962) 473–497, https://doi.org/10.1111/j.1399-3054.1962.tb08052.x.
[75] P. Klíma, V. Čermák, M. Srba, K. Müller, J. Petrášek, J. Šonka, L. Fischer, Z. Opatrný, Plant cell lines in cell morphogenesis research: from phenotyping to -omics, in: F. Cvrčková, V. Žárský (Eds.), Plant Cell Morphog. Methods Protoc, Springer New York, New York, NY, 2019, pp. 367–376, https://doi.org/10.1007/978-1-4939-9469-4_25.
[76] J.L. White, J.M. Kaper, A simple method for detection of viral satellite RNAs in small plant tissue samples, J. Virol. Methods 23 (1989) 83–93, https://doi.org/10.1016/0166-0934(89)90122-5.
[77] S.A. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele, C.T. Wittwer, The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, Clin. Chem. 55 (2009) 611–622, https://doi.org/10.1373/clinchem.2008.112797.
[78] C. Ramakers, J.M. Ruijter, R.H.L. Deprez, A.F.M. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, Neurosci. Lett. 339 (2003) 62–66, https://doi.org/10.1016/S0304-3940(02)01423-4.
[79] G.W. Schmidt, S.K. Delaney, Stable internal reference genes for normalization of real-time RT-PCR in tobacco (Nicotiana tabacum) during development and abiotic stress, Mol. Genet. Genomics 283 (2010) 233–241, https://doi.org/10.1007/s00438-010-0511-1.
[80] D. Tyč, E. Nocarová, L. Sikorová, L. Fischer, 5-Azacytidine mediated reactivation of silenced transgenes in potato (Solanum tuberosum) at the whole plant level, Plant Cell Rep. 36 (2017) 1311–1322, https://doi.org/10.1007/s00299-017-2155-7.
[81] D.M. Bond, D.C. Baulcombe, Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in Arabidopsis thaliana, Proc. Natl. Acad. Sci. 201413053 (2015), https://doi.org/10.1073/pnas.1413053112.
[82] T. Fehlmann, S. Reinheimer, C. Geng, X. Su, S. Drmanac, A. Alexeev, C. Zhang, C. Backes, N. Ludwig, M. Hart, D. An, Z. Zhu, C. Xu, A. Chen, M. Ni, J. Liu, Y. Li, M. Poulter, Y. Li, C. Stähler, R. Drmanac, X. Xu, E. Meese, A. Keller, cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs, Clin. Epigenetics 8 (2016) 123, https://doi.org/10.1186/s13148-016-0287-1.
[83] E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, B.A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, D. Blankenberg, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, Nucleic Acids Res. 46 (2018) W537–W544, https://doi.org/10.1093/nar/gky379.
[84] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (2015) 357–360, https://doi.org/10.1038/nmeth.3317.
[85] H. Guilley, R.K. Dudley, G. Jonard, E. Balàzs, K.E. Richards, Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts, Cell. 30 (1982) 763–773, https://doi.org/10.1016/0092-8674(82)90281-1.
[86] M. Bevan, W.M. Barnes, M.D. Chilton, Structure and transcription of the nopaline synthase gene region of T-DNA, Nucleic Acids Res. 11 (1983) 369–385, https://doi.org/10.1093/nar/11.2.369.
[87] G. Ji, L. Li, Q.Q. Li, X. Wu, J. Fu, G. Chen, X. Wu, PASPA: a web server for mRNA poly(A) site predictions in plants and algae, Bioinformatics 31 (2015) 1671–1673, https://doi.org/10.1093/bioinformatics/btv004.