A conserved Pol II elongator SPT6L mediates Pol V transcription elongation to regulate RNA-directed DNA methylation in *Arabidopsis*

Yujuan Liu[1,7], Jie Shu[1,2], Zhi Zhang[1,7], Ning Ding[1,8], Jinyuan Liu[1,7], Jun Liu[2], Yuhai Cui[3,6], Changhu Wang[1,7*], Chen Chen[1,4,5,7*]

[1]Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement & Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, 510650, China

[2]Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, 510640, China

[3]Agriculture and Agri-Food Canada, London Research and Development Centre, London, Ontario, N5V 4T3, Canada

[4]State Key Laboratory of Plant Diversity and Specialty Crops, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China

[5]Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China

[6]Department of Biology, Western University, London, Ontario, N6A 5B7, Canada

[7]University of the Chinese Academy of Sciences, Beijing, 100049, China

[8]MOE Key Laboratory of Cell Activities and Stress Adaptations, School of Life Sciences, Lanzhou University, Lanzhou 730000, China

[*]To whom correspondence should be addressed. Tel: +86 020-37252711; Email: chenchen101@scbg.ac.cn (to Chen Chen) and wangchangh@scib.ac.cn (to Changhu Wang)

Abstract

In plants, the plant-specific RNA polymerase V (Pol V) transcripts non-coding RNAs and provides a docking platform for the association of accessory proteins in the RNA-directed DNA methylation (RdDM) pathway. Various components have been uncovered that are involved in the process of DNA methylation, but it is still not clear how the transcription of Pol V is regulated. Here, we found that the conserved Pol II elongator, SPT6L, bound to thousands of intergenic regions in an RNA polymerase II (Pol II) independent manner. The intergenic enrichment of SPT6L, interestingly, co-occupied with the largest subunit of Pol V (NRPE1) and mutation of SPT6L led to the reduction of DNA methylation but not Pol V enrichment. Furthermore, the association of SPT6L at Pol V loci was dependent on the Pol V associated factor, SPT5L, rather than the presence of Pol V, and the interaction between SPT6L and NRPE1 was compromised in *spt5l*. Finally, Pol V RIP-seq revealed that SPT6L is required to maintain the amount and length of Pol V transcripts. Our findings thus uncovered the critical role of a Pol II conserved elongator in Pol V mediated DNA methylation and transcription, and shed light on the mutual regulation between Pol V and II in plants.

Introduction

In eukaryotic cells, transcription elongation is a dynamic and highly regulated process, in which a variety of functionally distinct transcript elongation factors are involved in Pol II progression[1,2]. Among them, the conserved elongator, SPT6, is recruited by the phosphorylated Pol II[3] and involved in the enhancement of elongation rate[4-6], repression of intragenic initiation[7,8], and transcription termination[5] in yeast and animal cells. In plants, the functional homolog of SPT6, SPT6-like (SPT6L), interacts with phosphorylated Pol II and plays conserved roles in Pol II progression[9]. The mutation of SPT6L causes pleiotropic defects in embryogensis[10] and post-germination stages[9]. Recently, it was found that SPT6L was able to recruit chromatin remodelers SWI2/SNF2 at transcription start sites (TSS) in a Pol II independent manner[11], indicating a potential role of SPT6L in transcription initiation/early elongation in plants.

Different from animal and yeast cells, in plants, two plant-specific RNA polymerases (Pol IV and V) have evolved and they play essential roles in the establishment and

maintenance of DNA methylation through the RdDM pathway[12,13]. In general, the canonical RdDM pathway is composed of two parts: the production of 24-nt siRNA and the establishment of DNA methylation[13]. The production of 24-nt siRNA is accomplished by Pol IV's transcription, RNA-DEPENDENT RNA POLYMERASE2 (RDR2)'s generation of double-stranded RNA, and DICER-LIKE PROTEINs (DCLs) dependent cleavage[13]. In the second part, Pol V transcripts serve as a docking platform to recruit AGOs-siRNA complex and other accessory proteins to establish DNA methylation. The RdDM pathway is a self-reinforcing loop[14] and the reduced siRNA and DNA methylation negatively affect the transcription of Pol V[12,13,15].

Unlike the processive transcription of Pol II, the estimated transcripts of Pol IV and V are short (30 to 40 nt[16,17] and around 200 nt[18], respectively) in length. Previous *in vitro* data indicates that both Pol IV and V can transcript on bipartite DNA-RNA templates and only Pol IV maintains the transcription ability on tripartite template[19], suggesting Pol V prefers to single strand DNA as template and is lack of ability to displace the non-template DNA. Recent structural data also shows that the conserved tyrosine residue of NRPE(D)2, the second subunit of Pol IV and V, can stall transcription and enhance backtracking by interacting with non-template DNA strand[20]. In addition, the lack of surfaces to recuit Pol II transcription factors such as TFIIB and TFIIS[17,20] also contributes to the inefficient transcrpiton of Pol IV and V *in vivo*. Although the above *in vitro* and structural data also revealed the nature of Pol IV and V in transcription, it is still not clear how the transcription of Pol IV and V are regulated in vivo and what factors are involved in the above process to distinguish the different transcription behaviors of Pol IV and V.

In this work, we found that the conserved elongator, SPT6L, was enriched at thousands of intergenic regions in a Pol II-independent manner. Interestingly, NRPE1, the largest subunit of Pol V, was also highly enriched in those regions. Mutation of SPT6L led to the reduction of DNA methylation but not the association between Pol V and chromatin. Further analyses showed that the associated protein, SPT5L, rather than the presence of Pol V is indispensable for the intergenic enrichment of SPT6L, and the interaction between SPT6L and NRPE1 was compromised in *spt5l*. Finally, NRPE1 RIP-seq indicated that SPT6L is required to maintain the amount and length of Pol V transcripts.

Taken together, our work revealed a Pol II and V shared component and its roles in the maintenance and promotion of DNA methylation and Pol V transcripts, respectively.

Results

SPT6L associates and co-occupies with the Pol V complex

Our previous data revealed that the elongation factor, SPT6L, associates with Pol II and plays roles in transcription initiation and elongation[9,11]. When browsing the occupancy signals of SPT6L, we have noticed the intergenic enrichment of SPT6L (Figure 1A). To examine the intergenic enrichment of SPT6L in detail, we reanalyzed previous ChIP-seq data[11] and identified 2,325 intergenic peaks across the genome (Supplementary Dataset 1). Within those regions, interestingly, we only detected the enrichment of SPT6L but not Pol II (Supplementary Figure 1A), which recruited SPT6L during the transcription of protein-coding genes[9]. We further analyzed the overlapping of the intergenic peaks with different genome features and found that transposons were highly enriched within those peaks (Supplementary Figure 1B), suggesting a potential link between SPT6L and the regulation of transposon elements (TEs).

The unexpected enrichment of SPT6L at transposons (Supplementary Figure 1B) and its conserved roles in DNA-dependent RNA polymerases[9,21,22] prompted us to examine the potential interplay between SPT6L and Pol IV/V, which play major roles in the silencing of transposons[12]. We profiled the published ChIP-seq signals of NRPE1[23], the large subunits of Pol V, at SPT6L binding sites and found highly enriched NRPE1 signals at the intergenic peaks of SPT6L (Figure 1A and Supplementary Figure 1A). By comparing the binding peaks of SPT6L and NRPE1, we found 6,008 NRPE1 peaks overlapped with SPT6L peaks (Supplementary dataset 1) and a stronger binding strength of NRPE1 at the overlapped peaks than that at NRPE1-only peaks (Supplementary Figure 1C). We further compared the frequency of TE within NRPE1 peaks and observed a higher amount of TE in NRPE1-SPT6L overlapped peaks than that in NRPE1-only peaks (Supplementary Figure 1D). To clarify the types of TE associated to NRPE1 and NRPE1-SPT6L peaks, we found that the NRPE1-bound TEs were more abundant in Helitron and SINE, but less enriched in Gypsy (Supplementary Figure 1E). The compositions of TE within NRPE1

and NRPE1-SPT6L peaks were similar (Supplementary Figure 1E), indicating that there is no preference of NRPE1-SPT6L peaks in different TE types.

As the published ChIP-seq data of NRPE1 was sourced from inflorescence[23], we decided to generate GFP-tagged NRPE1 transgenic lines and profiled the genome-wide occupancy of NRPE1 with the same plant tissues as those used for generating the SPT6L's data. Firstly, we confirmed the normal function of NRPE1-GFP (*nrpe1-11 pNRPE1:NRPE1-GFP*, hereafter all *nrpe1-11* were named *nrpe1*) by examining the nuclear-localized GFP signals (Supplementary Figure 2A) and recovered DNA methylation at selected RdDM loci (Supplementary Figure 2B). And then, we profiled the genome-wide occupancy of NRPE1 and identified 7,809 confident peaks (Irreproducible Discovery Rate, IDR < 0.01, Supplementary dataset 1) across two biological replicates. Those peaks were highly overlapped with published NRPE1 binding peaks (Supplementary Figure 2C). By comparing the binding signals of NRPE1 and SPT6L, consistently, we detected similar co-binding signals between SPT6L and NRPE1 at SPT6L intergenic peaks, where were lack of Pol II signals (Figure 1B). Finally, to define SPT6L and NRPE1 co-bound genomic regions in a reliable and unbiased way, we divided the genome into 200 bp bins and used a Hidden Markov Model (HMM) to identify bins with enrichment of SPT6L and NRPE1. When the normalized SPT6L, NRPE1, and Pol II ChIP-seq data combined from two biological replicates were analyzed this way, the genome could be split into six groups (Supplementary Dataset 2) and, importantly, the NRPE1-only (G2, weak SPT6L signals) and NRPE1-SPT6L shared (G3, strong SPT6L signals) bins (Figure 1C and Supplementary Figure 2D) were clearly distinguished. Consistently, the NRPE1-SPT6L shared regions contained more TE loci than the NRPE1-only ones (Supplementary Figure 2E). Although the NRPE1 signals were generally enriched upstream of transcription start sites (TSS), the NRPE1-SPT6L shared and NRPE1-only bins were, interestingly, distinguished around 400 (distal) and 200 (proximal) bp upstream of TSS, respectively (Figure 1D), suggesting that these two different binding patterns of NRPE1 may have distinct roles.

To examine the potential association between SPT6L and Pol V complex, we performed yeast-two hybrid assays between SPT6L and multiple subunits, which were distinct

between Pol II and V (NRPE1, NRPE2, NRPE4, NRPE5, and NRPE7)[24]. As shown in Supplementary Figure 2F, no interaction was found between SPT6L and NRPE1, while SPT6L can directly interact to NRPE4 and NRPE7 with different strength. And then, we further confirmed the observed interactions among SPT6L and several subunits of Pol V by co-immunoprecipitation (Co-IP) with stable transgenic lines (Figure 1E; Supplementary Figure 2A and 2G). Consistently, we were also able to detect the associations between SPT6L and multiple subunits of Pol V in *vivo* (Figure 1E), indicating that SPT6L can form protein complex with Pol V in *planta*. Altogether, the above results indicate that SPT6L probably collaborates with Pol V to mediate the silencing of transposons in plants.

Pol V is required for the intergenic recruitment of SPT6L

The interaction between, and genomic co-occupancy of SPT6L and Pol V prompted us to examine the mutual dependency of their genome recruitment. Therefore, we profiled and compared the genome-wide occupancy of NRPE1 in WT (*nrpe1 pNRPE1:NRPE1-GFP*) and *spt6l* (*nrpe1 spt6l pNRPE1:NRPE1-GFP*) backgrounds and found that the overall enrichment of NRPE1 in *spt6l* was unchanged in both G3 (NRPE1-SPT6L shared) and G2 (NRPE1-only) regions (Figure 2A). The following ChIP-qPCR at selected RdDM loci also confirmed the general unchanged pattern of NRPE1 occupancy in *spt6l* (Supplementary Figure 3A). We further assessed the protein stability of NRPE1 in *spt6l* and detected a comparable protein level of NRPE1 in both WT and *spt6l* (Supplementary Figure 3B). By comparing the genome-wide profiles of NRPE1 in WT and *spt6l*, we found that the ChIP reads of NRPE1 in both genotypes were highly correlated (Figure 2B). These results indicate that SPT6L is dispensable for the enrichment of NRPE1. Next, we examined the dependency of SPT6L on NRPE1 by ChIP-seq in WT and *nrpe1* mutant backgrounds. As shown in Figure 2C, the mutation of *NRPE1* dramatically reduced the occupancy of SPT6L at G3 (NRPE1-SPT6L shared) regions but not other SPT6L enriched regions (G4 and G5). These results were then confirmed by ChIP-qPCR at selected loci (Supplementary Figure 3C) and an immunoblotting showed a comparable protein level of SPT6L in both WT and *nrpe1*(Supplementary Figure 3D). This result indicates that NRPE1 is required for the intergenic enrichment of SPT6L in plants.

Although the dependency of SPT6L on NRPE1 shown above explained the co-occupancy of SPT6L and NRPE1 at the G3 regions, it is still not clear why there were less enriched SPT6L signals at the G2 regions, which showed a moderate NRPE1 signal (Figure 1C). We noticed that the G2 regions were enriched around 200 bp upstream of TSS, which is closer than that of G3 regions to TSS (Figure 1D). As there is strong association of SPT6L with phosphorylated Pol II around TSS[9], we assumed that the weak signals of SPT6L at the G2 regions may result from the local competition between Pol II and V in the recruitment of SPT6L. To test this assumption, firstly, we identified the nearest downstream TSS of each previously defined genomic group (G2 to G5) (Supplementary Dataset 3) and compared the SPT6L ChIP signals around those TSS. As shown in Figure 2D, the mutation of *NRPE1* led to a slight but clear increase of SPT6L occupancy only at the downstream genes of the G2 regions, suggesting that the presence of upstream NRPE1 in the G2 regions may either trap SPT6L or directly inhibit transcription. And then, we compared the SPT6L ChIP signals[9] at four different groups after treating with a P-TEFb inhibitor (Flavopiridol, FLA), which decreases the phosphorylation levels of Pol II and disrupts its interaction with SPT6L[9,11]. Indeed, application of the inhibitor reduced the occupancy of SPT6L at genic regions (G4 and G5) (Figure 2E and Supplementary Figure 2E). Meanwhile, we also detected increased occupancies of SPT6L at both G2 and G3 regions (Figure 2E), indicating that the dissociation of SPT6L to Pol II can increase the occupancy of SPT6L at NRPE1 binding sites. Altogether, the above results suggested that Pol II and V may compete to recruit SPT6L to facilitate its transcription in plants.

Mutation of SPT6L causes the reduction of DNA methylation

The essential role of Pol V in RdDM and the intergenic enrichment of SPT6L led us to further examine the potential effects of SPT6L on DNA methylation in plants. We first performed Chop-PCR to examine the DNA methylation at several known RdDM loci. As shown in Figure 3A, the DNA methylation levels were reduced but not eliminated at the *SN1*, *IGN5*, and *IGN23* loci in *spt6l*. And then, to assess the generality of these findings, we performed whole-genome bisulfite sequencing analyses (BS-seq) and identified 4,099 differentially methylated regions (DMRs) in *spt6l* (Supplementary Dataset 3). Most of the DMRs (3,681 out of 4,099) were hypomethylated. Similar to what was found in *nrpe1*,

the identified hypo DMRs showed hypomethylation at the CHG and CHH contexts (hypo mCHG and mCHH) compared to WT (Figure 3B). We then performed a BS-seq in the *nrpe1 spt6l* double mutant and revealed a similar hypo mCHG and mCHH to that in *nrpe1* (Figure 3B), indicating that the mCHG/mCHH in hypo DMRs of *spt6l* mainly NRPE1-depedent, and SPT6L may be involved in mCHG/mCHH through NRPE1. Interestingly, the mutation of SPT6L also caused a dramatical reduction of DNA methylation at the CG context (mCG), which was only slightly reduced in *nrpe1*(Figure 3B), indicating that SPT6L may also regulate DNA methylation in NRPE1 independent manner. To examine whether SPT6L directly contributed to the DNA methylation at the hypo DMRs in *spt6l*, we integrated our ChIP-seq and BS-seq data and found that the decreased methylation, even in the CG context, was mainly detected at G2 and G3 regions (Figure 3C and 3D), suggesting that SPT6L likely contributes to DNA methylation mainly through NRPE1 mediated DNA methylation.

As SPT6L interacts with Pol II and plays an essential role in transcription[9,21], the reduced DNA methylation may result from the mis-expressed genes encoding DNA methyltransferases and demethylases in *spt6l*. Thus, to assess the possibility, we performed RNA-seq assay with three biological replicates and compared the expression of DNA methyltransferases and demethylases in WT and *spt6l*. In total, we have detected more than 12,000 differentially expressed genes in *spt6l* with DEseq2 package (Supplementary Figure 4A and Supplementary Datasets 4, adjust $P < 0.01$, |Fold Change| ≥ 2). Within the five major DNA methyltransferase genes (*MET1*, *CMT2*, *CMT3*, *DRM1*, and *DRM2*)[25], we found that only the expression of *CMT2* was significant down-regulated in *spt6l* (Figure 3E). Except for *CMT2*, the *DRM1* and the other three genes encoding DNA methyltransferase showed increased and unchanged expression in *spt6l* mutant, respectively (Figure 3E). Meanwhile, we also examined the expression of four major DNA demethylase genes (*ROS1*, *DME*, *DML2*, and *DML3*) and detected a significant decrease of *ROS1* and *DML2* in *spt6l* (Figure 3E). To further estimate the potential effects of malfunctioned transcription on DNA methylation, we took advantage of the published DNA methylation datasets[26] and compared the methylation levels at *spt6l* DMRs in mutants of five methyltransferases, Pol IV (*nrpd1*), and Pol II (*nrpb2*). As

shown in Supplementary Figure 4B, the changed DNA methylation patterns at the three contexts in *spt6l* were distinct from that in *cmt2*, which caused changed DNA methylation mainly at CHG/CHH. Altogether, the above results suggest that the reduced DNA methylation in *spt6l* is less likely resulted from the mis-regulation of DNA methyltransferases and demethylases.

Other than DNA methyltransferases, the biogenesis of small interfering RNA (siRNA) also plays an essential role in both the canonical and non-canonical RdDM pathways in plants[13]. Thus, we examined the expression of multiple components related to the production of siRNA and found the expressions of *NRPD1* and *DCL3* were altered in *spt6l* (Figure 3E). To directly estimate the potential effect of SPT6L on the production of siRNA, we performed small-RNA deep sequencing in WT and *spt6l* for comparison. Even though some of the 21-22 and 24-nt siRNA produced by Pol II[27], interestingly, we did not detect any dramatic change in the compositions of the 21-22 and 24-nt siRNAs in *spt6l* (Figure 3F). Previously, the 24-nt siRNAs have been clustered into upstream (siRNAs dependent on Pol IV only) and downstream siRNAs (siRNAs dependent on both Pol IV and Pol V)[28]. The upstream siRNAs are affected only in mutants defective in upstream RdDM components, such as *nrpd1*, whereas the downstream siRNAs are affected in the mutants of both Pol IV and V related components. To carefully assess the role of SPT6L in the biogenesis of siRNA, we compared the amounts of 24-nt siRNA in the above two clusters between WT and *spt6l*. As shown in Figure 3G, a slightly decreased amount of the 24-nt siRNA was found in Pol V dependent regions, although the total composition of 24-nt siRNA was not reduced in *spt6l* (Figure 3F). Meanwhile, we also detected an unchanged or even slightly increased siRNA in Pol IV only regions (Figure 3G), which may result from the up-regulation of *NRPD1* in *spt6l* (Figure 3E). These results indicated that the production of siRNA is generally not reduced in *spt6l* mutant and the reduced DNA methylation in *spt6l* unlikely results from the alternation of siRNA.

The WG/GW repeat of SPT6L is dispensable for RdDM

The C-terminals of both NRPE1 and SPT5L contain a WG/GW repeat, which is essential for the AGO4 recruitment and DNA methylation at RdDM loci[29]. Interestingly, a

WG/GW repeat was also found at the C-terminal of SPT6L, which was computationally scored in the top 3 of Argonaute (AGO) interacting proteins[30]. To examine whether the WG/GW repeat of SPT6L contributes to the enrichment of SPT6L at RdDM loci and DNA methylation, we generated a WG/GW deleted construct and introduced it into *spt6l*$^{+/-}$. As shown in Figure 4A, the truncated SPT6L was able to rescue the developmental defects of *spt6l*. The transgenic line of *spt6l SPT6LΔWG/GW-GFP* was further validated by confirming its nuclear localization signals and comparable protein levels to that of SPT6L (Supplementary Figure 5A and 5B). And then, the genome-wide occupancy of SPT6LΔWG/GW-GFP was profiled and a similar binding pattern and high correlation were revealed between SPT6LΔWG/GW and SPT6L (Figure 4B and 4C), suggesting that the WG/GW repeat is not required for the transcriptional function of SPT6L under normal condition. Especially, the occupancy of SPT6LΔWG/GW at the SPT6L intergenic regions was comparable to that in SPT6L (Figure 4B), indicating that the WG/GW repeat is also dispensable for the association of SPT6L to RdDM loci. To examine the role of SPT6L-WG/GW in DNA methylation, we performed Chop-PCR and found that the introduction of SPT6LΔWG/GW was able to recover the reduced DNA methylation at selected RdDM loci (Figure 3A). We then performed BS-seq and detected similar genome-wide DNA methylation levels between WT and *spt6l* SPT6LΔWG/GW (Figure 4D and 4E). Altogether, these results indicate that the WG/GW repeat of SPT6L, unlike that of NRPE1 and SPT5L, is dispensable for SPT6L's genomic recruitment and role in DNA methylation at the RdDM loci.

SPT5L is required for the recruitment of SPT6L to RdDM loci

In the RdDM pathway, following the recruitment of Pol V, multiple proteins are bound to Pol V/Pol V transcripts and mediate the DNA methylation[13]. To further clarify whether the enrichment of SPT6L at intergenic loci is dependent on Pol V or the downstream events, we examined the genome-wide occupancy of SPT6L in *spt5l*, which impairs the slicing features of Pol V transcripts but not the enrichment of Pol V [31-33]. Interestingly, the occupancies of SPT6L at the NRPE1 related regions (G2 and G3) but not the other SPT6L enriched regions (G4 and G5) were dramatically reduced in *spt5l* (Figure 5A and 5B), indicating that SPT5L is required for the intergenic recruitment of SPT6L. We

confirmed the results at selected genomic loci by ChIP-qPCR (Supplementary Figure 6A), and our immunoblotting assay showed that the altered enrichment of SPT6L did not result from any potential changes of protein stability in *spt5l* (Supplementary Figure 6B). To confirm the potential effects of *spt5l* on the binding of NRPE1, we examined the enrichment of NRPE1 at selected loci in *spt5l* by ChIP-qPCR. Consistent with previous results[33], the binding of NRPE1 was generally unchanged in *spt5l* at selected loci (Figure 5C). These data indicated that the presence of Pol V alone is insufficient to determine the binding of SPT6L. To further examine the essential role of SPT5L in the recruitment of SPT6L, we firstly examined the interaction between them by yeast-two hybrid assays. As shown in Supplementary 6C, an interaction between SPT5L and SPT6L was detected in yeast and the further truncations of SPT5L revealed the N-terminal of SPT5L played a major role in its interaction with SPT6L. To confirm the interaction *in vivo*, stable transgenic plants containing *pSPT5L:SPT5L-GFP* and *pSPT6L:SPT6L-MYC* were generated and the interaction between SPT5L and SPT6L was confirmed by Co-IP assay (Figure 5D). Finally, by knocking out SPT5L, we detected a compromised interaction between SPT6L and NRPE1 (Figure 5E), indicating that SPT5L is indispensable for the recruitment of SPT6L into the Pol V complex.

In addition, we also examined the occupancy of SPT6L in the mutants of *drm1 drm2* and *nrpd1*. The former plays a role in the downstream of Pol V and catalyzes DNA methylation[14]. The latter encodes the largest subunit of Pol IV and determines the production of 24-nt siRNA in plants[34]. As shown in Figure 5F, the occupancies of SPT6L in *drm1 drm2* and *nrpd1* are significant reduced at some but not all selected loci, suggesting that Pol IV and DRM1/DRM2 may affect the intergenic enrichment of SPT6L in a loci-specific manner. Immunoblotting confirmed that the changed occupancies of SPT6L in *drm1 drm2* and *nrpd1* did not result from protein stability (Supplementary Figure 6D). Previously, it was reported that the occupancy of Pol V in *nrpd1* and *drm1 drm2* were slightly reduced[23,32,35]. Thus, the reduced occupancies of SPT6L in *nrpd1* and *drm1 drm2* may result from the decreased occupancy of Pol V.

SPT6L is required for Pol V elongation

In eukaryotic cells, SPT6L(SPT6) plays an essential role in Pol II elongation[5,9]. The association of SPT6L with Pol V led us to examine the potential functions of SPT6L in Pol V elongation. For that, we performed RIP-seq in *nrpe1*, *NRPE1-GFP nrpe1*(*NRPE1-GFP*), and *NRPE1-GFP nrpe1 spt6l* (*spt6l NRPE1-GFP*) by using a GFP antibody. To minimize the effect of mechanical force on Pol V transcripts, we replaced the sonication step with DNase I treatment in the original IPARE protocol[31]. As shown in Figure 6A to 6C, the Pol V transcripts can be detected in NRPE1 peaks and the amount of Pol V transcripts within NRPE1 peaks was significantly reduced in *spt6l*, indicating that SPT6L is required for promoting Pol V transcription. Furthermore, by comparing the RIP-seq reads, we also found a significant reduction of the length of RIP-seq reads in *spt6l* (Figure 6D), suggesting that SPT6L may play a role in Pol V elongation. Other than the quantity and length of RIP-seq reads, another feature of Pol V transcription worth examining is that Pol V transcripts share a special slicing feature at the +10 position, which is dependent on SPT5L[32]. To examine whether the slicing feature was also affected in *spt6l*, we examined the base frequency around the 5' end of RIP-seq reads. Interestingly, in *NRPE1-GFP* and *spt6l NRPE1-GFP*, we both detected a slightly purines (G+A) preference at the initiation site (+1), but not the +10 U preference which was revealed previously[32]. This inconsistency likely results from the different strategies used for library preparation. Altogether, these results indicate that SPT6L is required for sustaining and promoting the transcription of Pol V.

Discussion

The two plant-specific RNA polymerases Pol IV and V play essential roles in the RdDM pathway. Many accessories of these two polymerases were successfully identified in the last two decades, but it is still not clear how the transcription process of Pol IV and V are regulated *in vivo*. In this work, we reported the physical association of a conserved Pol II elongator, SPT6L, with the Pol V complex and investigated the roles of SPT6L in the regulation of DNA methylation and Pol V transcription. Our findings indicate a conserved transcription regulation mechanism between these two transcription complexes. Although this is not totally surprising as several Pol II and V shared factors such as

AGO4[27], RDM1[36], and RDM4[37] have been identified, it is rather exciting in that SPT6L is the first elongation factor found to play such a important role.

Pol V dependent DNA methylation serves as the main mechanism to repress the transcription of both TEs and downstream genes[12]. Indeed, knocking out NRPE1 resulted in much marked up-regulation of Pol V-proximal genes[38]. Interestingly, the co-occupancies of SPT6L and NRPE1 were mainly detected at TSS-distal (-600 to -200 bp upstream of TSS) rather than TSS-proximal (-200 bp to TSS) regions (Figure 1D). By knocking out NRPE1 and blocking the SPT6L-Pol II interaction, we found an increased enrichment of SPT6L at the nearest downstream TSS of NRPE1 (Figure 2D) and NRPE1 binding sites (Figure 2E), respectively. These results suggest that Pol II may directly compete with Pol V in the recruitment of SPT6L and then lead to the low enrichment of SPT6L at TSS-proximal NRPE1 loci. Future works are needed to test the potentially mutual regulation between Pol II and V in the competition for core transcription accessories.

Loss of Pol V mainly causes the reduction of DNA methylation at CHG and CHH[38]. While the requirement for Pol V on the intergenic enrichment of SPT6L (Figure 2C), the reductions of DNA methylation in *spt6l* were detected in all three contexts (CG, CHG, and CHH) (Figure 3B to 3D). According to the amounts of siRNAs, the reduction of DNA methylation in *spt6l* is unlikely resulted from alteration of siRNA production. The general reduction of mCHG and mCHH is partially contributed by the down-regulation of *CMT2* (Figure 3E) and mis-regulated Pol V transcripts (Figure 6A to 6D). Referring to the decreased mCG, we found the reduced mCG was mainly detected in the NRPE1 bound regions (Figure 3D) and the decreased mCG in *spt6l* was partially recovered in *spt6l nrpe1* (Figure 3B and 3D), suggesting a negative effect of NRPE1 in *spt6l* on the level of mCG. As the binding profile of NRPE1 was unaffected in *spt6l*, the occupancy of NRPE1 may block the access of CG methyltransferases such as MET1. In the future, it will be interesting to examine how the different types of DNA methylation are mutually affected to each other.

The SPT6L was computationally characterized as one of the top 3 proteins that contained WG/GW repeats[30], a well-known domain to interact with AGOs[30]. However, the SPT6L-

WG/GW contributed to neither the intergenic enrichment of SPT6L nor the DNA methylation at RdDM loci (Figure 4B and 4F), suggesting that this repeat may be dispensable in DNA methylation. The simultaneous deletion of WG/GW repeats both in NRPE1 and SPT5L reduces the level of AGO4 enrichment and DNA methylation to that of in *nrpe1-11*[29], suggesting that the presence of SPT6L-WG/GW has little contribution to the recruitment of AGO4 and the DNA methylation at RdDM loci.

In the RdDM pathway, Pol V and its transcripts provide a docking platform for downstream components[39]. The mutation of SPT5L dramatically reduced the intergenic enrichment of SPT6L (Figure 5B) and compromised the association between SPT6L and NRPE1 (Figure 5E), suggesting that Pol V downstream events rather than Pol V itself determines the intergenic recruitment of SPT6L. The Pol V complex with SPT5L being recruited to it may represent an active state of Pol V, which can further recruit other accessory components such as SPT6L. SPT5L is a homolog of SPT5, which physically contacts SPT6 through its KOW domain in animal cells[22]. In line with this association, a physical interaction between SPT6L and SPT5L was also detected in its N-terminal, which contains the KOW domain (Supplementary Figure 6C). Interestingly, SPT4, another interacting partner of SPT5, is also involved in the regulation of DNA methylation[40], suggesting that SPT6L may not be the only Pol II and V shared elongators.

Our NRPE1 RIP-seq identified the weak enrichment of Pol V transcript 5’ ends at purines (A/G) (Figure 6E), but the potential bias of template-switching in library preparation may also contribute to this feature[41]. Thus, cautions need to be taken when drew conclusion about the 5’-end feature of Pol V transcripts. In addition, we did not detect the U preference at +10 of Pol V transcripts, which was revealed previously in another study through GRO-seq[32]. This inconsistency likely results from the different strategies used in library preparation. In the previous GRO-seq, 5’ monophosphorylated (5’-p) RNAs were selectively enriched for library preparation[32], while we generated the Pol V RIP library by template switching, which was able to rescue multiple 5’end of RNA such as 7-methylguanosine capped, 5’ phosphates, and 5’ hydroxyl RNAs[41]. Based on the RNA levels of serveal *IGN* loci, previous work estimated that about 70% of Pol V transcripts

with 5'-triphosphate (5'-ppp) end and 30% transcirpts with 5'-p[42]. Thus, the inconsistent feature may represent different states of Pol V transcripts.

The role of SPT5L in Pol V transcription is still in debate. Previous RT-PCR[43] and IPARE[31] data showed unchanged Pol V transcripts in *spt5l*. The GRO-seq results[32], on the other hand, revealed the roles of SPT5L both in slicing and the amount of Pol V transcripts. In this work, we show that SPT5L plays an essential role in determining the intergenic association of SPT6L (Figure 5B), and the mutation of SPT6L reduced the amount and length of Pol V transcripts (Figure 6A to 6D), suggesting that SPT5L may be involved in the regulation of Pol V transcripts. The inconsistency may due to our modification of the original IPARE protocol, in which the sonication step was replaced by DNase I treatment. Future works may need to clarify the role of SPT5L in Pol V transcription.

Methods

Plant Materials and Growth

Arabidopsis seeds were stratified for 2 d at 4 °C in darkness. The seeds were then sown on soil or agar plates containing 2.22 g/L Murashige and Skoog (MS) nutrient mix (PhytoTech LABS, M519), 1.5% sucrose (pH 5.8), and 0.8% agar. Plants were grown in growth rooms with 16-h light/8-h dark cycles at 22 °C. All Arabidopsis lines used in this study were in Columbia (Col-0) background. The mutants of *nrpd1-3* (SALK_083051)[34], *nrpe1-11* (SALK_029919)[44], *drm1/2* (CS16383)[45], and *spt6l* (SALK_016621)[9,10] were previously reported. The seeds of *spt5l-1* (SALK_001254C)[46] and *ProNRPE1:NRPE1-FLAG* (CS66156)[47] were obtained from the Arabidopsis Biological Resource Center (ABRC). The seedlings of *spt6l* and *nrpe1-11 spt6l* homozygous were respectively selected from *spt6l*$^{+/-}$ and *nrpe1-11 spt6l* $^{+/-}$progenies based on its defected phenotypes as reported previously[9]. To obtain the *nrpe1 spt6l NRPE1-GFP* seedlings, we transformed the *ProNRPE1:NRPE1-GFP* construct into *nrpe1 spt6l*$^{+/-}$ plants and selected the correct seedlings from the progenies of homozygous *nrpe1 spt6l*$^{+/-}$ *ProNRPE1:NRPE1-GFP*. The transgenic lines *ProSPT6L:SPT6L-GFP* and *ProSPT6L:SPT6L-MYC* were previously reported[9,11]. The transgenic lines of *SPT6L-MYC NRPE1-GFP*, *SPT6L-MYC NRPE4-GFP*, and *SPT6L-MYC NRPE7-GFP* were generated by transforming

ProNRPE1:NRPE1-GFP, *ProNRPE4:NRPE4-GFP*, and *ProNRPE7:NRPE7-GFP* construct into *SPT6L-MYC* line. All materials used in this study were 10-day old otherwise specified elsewhere.

Plasmid construction

For the generation of transgenic plants, the full-length *NRPE1* and *SPT5L* genomic region and their ~2 kb upstream putative promoters were amplified and cloned into the *pMDC107*. Firstly, part1 (from 2,209 to 8,315 bp, relative to ATG) and part2 (from −2,088 to 2,220 bp) of *NRPE1*, part1 (from −2,124 to 3,675 bp) and part2 (from 3,655 to 6,582 bp) of *SPT5L* were PCR-amplified from genomic DNA. And then, both the part1 of *NRPE1* and *SPT5L* were inserted into *pMDC107* individually by using PmeI and AscI. Finally, the part2 of both genes were subcloned into *pMDC107-part1* by ClonExpress® II One Step Cloning Kit (Vazyme, C112). In addition, the genomic sequences contained upstream regulatory sequence of *NRPE4* (from -2,096 to 1,249 bp) and *NRPE7* (from -3,012 to 534 bp) were amplified and cloned into the *pMDC107*. For yeast two-hybrid assay, the CDS of *NRPE1*, *NRPE2*, *NRPE4*, *NRPE5*, and *NRPE7* were amplified and cloned into *pGADT7*. Truncated fragment of *SPT5L* CDS were amplified and cloned into *pGADT7* according to previous works[43,48]. The CDS sequence of *SPT6L* was amplified and cloned into *pGBKT7*. All primers used for plasmid construction are listed in Supplementary Table 1.

Y2H analysis

The vector for bait (*pGBKT7*) and prey (*pGADT7*) were co-transformed into yeast strain AH109 that was selected on medium lacking leucine (Leu) and tryptophan (Trp). Positive colonies were picked up and dropped on -Leu/-Trp/-His medium containing 10 mM E-amino-1, 2, 4 triazol (3-AT) for image recording.

Confocal microscopy

To detect green fluorescence signals, root tips were cut from 7-day-old seedlings and transferred onto glass slides with 50 μL H_2O. The green fluorescence was detected by confocal microscopy (Leica) with excitation at 488 nm and emission at 505 to 525 nm.

Immunoblotting and Co-immunoprecipitation

Two hundred milligrams of 10-day-old seedlings were harvested and homogenized to fine powder, which was subsequently dissolved in 300 μL lysis buffer (100 mM Tris–HCl pH 7.5, 300 mM NaCl, 2 mM EDTA, 0.5% TritonX-100, 10% glycerol, 1 mM PMSF, and protease inhibitor cocktail) for 30 min at 4°C with gentle shaking. Next, the crude lysate was centrifuged at 18,000 g for 10 min at 4°C to remove debris. For Western blot (WB), the supernatants were mixed with 4× SDS loading buffer and loaded onto SDS-PAGE gels. For Co-IP, we added 25 μL anti-GFP nanobody agarose beads (KT HEALTH, KTSM1301) to the supernatants and incubated for 2 h at 4°C with gentle shaking. The beads were washed five times with wash buffer (100 mM Tris-HCl (pH 7.5), 300 mM NaCl, 2 mM EDTA and 0.75 % TritonX-100). After centrifugation, the beads were boiled with 2× SDS sample buffer for 5 min. The interacting proteins then loaded onto SDS-PAGE gels. The antibodies used in this study are listed as follow: anti-GFP (Abcam, ab290; 1:20,000 dilution), anti-H3 (Abcam, ab1791; 1:20,000 dilution), anti-MYC (Abcam, ab9106; 1:20,000 dilution).

Chromatin immunoprecipitation and library preparation

For most of ChIP samples, ChIP assays were carried out as previously described[11]. For NRPE1-Flag ChIP, the nuclei were firstly enriched as previously described[49] and then followed with nuclei lysis. Immunoprecipitation was performed by using either anti-GFP antibody (Abcam, ab290) or anti-FLAG M2 Magnetic Beads (Sigma–Aldrich, M8823). For ChIP-qPCR, at least two biological replicates were included and primers were listed in Supplementary Table 1. For ChIP-seq, the libraries of ChIP DNA were prepared following the published protocol[50] with at least two biological replicates otherwise specified elsewhere. The reads information of different sample was collected in Supplementary Table 2. The correlations across biological replicates can be found in Supplementary Figure 7A

RNA and Small RNA-seq

Total RNA was extracted from 10-day-old seedlings using TRIzol (Invitrogen, 15596-018). Genomic DNA was removed by treating with TURBO DNase and then the DNase inactivated RNA was used for either mRNA or small RNA library preparation. For small RNA, RNA samples were separated on a PAGE gel, and the 18- to 30-nt fraction of the

gel was cut for small RNA purification. For RNA-seq, Poly(A) mRNAs was enriched with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Library preparation and sequencing were performed using Illumina reagents according to the manufacturer's instructions. The correlations across biological replicates can be found in Supplementary Figure 7B

Chop-PCR and whole-genome bisulfite sequencing

Genomic DNA was extracted from 10-day-old seedlings using DNeasy Plant Mini Kit (QIAGEN 69104). And then, about 100 ng genomic DNA was digested overnight with methylation-sensitive restriction endonucleases (HaeIII, NEB, R0108S). The digested DNA was used to amplify the indicated regions by PCR using primers flanking the endonuclease recognition sites. Primers are listed in Supplementary Table 1. For the bisulfite sequencing, the extracted DNA was directly sent to the NovoGene for whole-genome bisulfite sequencing (WGBS). The correlations across biological replicates can be found in Supplementary Figure 8

RNA immunoprecipitation and library preparation

The NRPE1 RIP-seq was performed as previously described[31,51,52] with modifications. Briefly, 2 g of 10-day-old seedlings was used for chromatin extraction. Chromatin was treated by DNase I (NEB, M0303S) for 1 hour, and then we added 1% final concentration SDS to treated Chromatin. Supernatant was diluted five times with chip diluent buffer. And IP was performed using 2 µL/IP anti-GFP (ab290, Abcam) at 4°C for overnight. After rescuing Pol V associated RNA, the removal of residual gDNA and addition of poly A tail were performed as described in IPARE protocol[31]. To increase the efficiency of reverse transcription 10 times higher amount of dCTP (0.5 µl 100 mM) was added into reverse transcription buffer[41]. The following DNA purification and library preparation were similar to that in IPARE protocol[31].

Bioinformatic analysis

ChIP-seq: The adaptors of raw ChIP-seq reads were removed by using cutadapt[53] (version 3.4, default settings) and mapped to Arabidopsis genome by Bowtie2[54] (version 2.4.2, default settings) in pair-end mode. The unmapped, improperly paired, and

duplicated reads were removed using samtools[55] (version 1.11, default settings). And then, bam files were converted to BEDPE format with bedtools (version 2.27.1, default settings) and the pair-end mode of MACS2[56] (2.2.7.1, -f BEDPE, -g 135000000, -q 0.001) was used to generate peak lists. The high confident peak list across two biological replicates was generated by using IDR (version 2.0.3, --idr-threshold 0.01). The bamCoverage of deeptools[57] (version 3.5.1, -bs 10, --normalizeUsing RPGC, --effectiveGenomeSize 135000000) was used to generate genome coverage files. The values under heatmaps and plots were generated with computeMatrix (subcommand of deeptools, -bs 10 –missingDataAsZero). The analyses of correlations between samples/replicates were performed by using multiBigwigSummary (subcommand of deeptools, BED-files mode, the regions of correlation analyses for NRPE1 and SPT6L IP samples were NRPE1 and SPT6L peaks, respectively). The different states of genome were identified by using ChromHMM[58] (version 1.24, default settings). The averaged coverage file from two biological replicates was generated by running a GitHub script (http://wresch.github.io/2014/01/31/mergebigwig-files.html). Genome tracks were generated with pyGenomeTracks[59].

BS-seq: The raw reads of BS-seq were processed by cutadapt to remove adaptors and aligned to Arabidopsis genome by using Bismark[60] packages (version 0.23.1, default settings). PCR duplicates were removed by using deduplicate_bismark (default settings). And then, the methylated bases were extracted by using bismark_methylation_extractor (default settings). Finally, the outputs of bismark_methylation_extractor can be load into a R package-methylkit[61] (version 1.22.0, mincov=4, win.size=500, step.size=500, difference=25, qvalue=0.01) to identify differential methylated regions and calculate correlation values.

RNA-seq: The cleaned RNA-seq raw reads were mapped into Arabidopsis genome with STAR (version 2.7.11a, default settings, --genomeSAindexNbases 12) and raw reads count per gene in each sample was calculated by RSEM package (version 1.3.3, with default settings). Finally, the analysis of differential expression was performed by using an R package-DEseq2 (version 3.18).

smRNA-seq: The reads quality of smRNA-seq was firstly checked by using FastQC (version 0.11.9) and the adaptor and linker were removed by using cutadapt. And then, the processed reads were mapped into *Arabidopsis* genome with Bowtie2 (default settings). After filtering out the unmapped reads, the mapped reads were converted to bed format with bedtools. Different sizes of small RNAs were selected, counted, and compared within different genome regions.

RIP-seq: The trimming, mapping, and removing PCR duplicates were performed as previously described[31] with minor modifications. Briefly, the 8 bp long unique molecular identifier (UMI) in the first read of paired-end reads were removed and appended to the read name by using UMI-tools[62] (version 1.1.2). And then, the reads of read 1 (first read in paried-end reads) were then trimmed to remove the 3'poly (A) and 5' TATAGGG (cutadapt, -m 10). Finally, the processed reads (read 1) were mapped Arabidopsis genome with default settings. The PCR duplicates were removed by using the UMI-tools and processed reads were converted to bed format with bedtools. To identify Pol V transcripts, the mapped RIP reads from NRPE1 or *spt6l* NRPE1 were intersected with reads identified in *nrpe1* and only the non-overlapped reads were kept. Finally, the processed RIP-seq reads were overlapped with NRPE1 peaks and only the reads within NRPE1 peaks were considered as Pol V transcripts.

Data availability

The data that support the findings of this study are available from the corresponding author upon request. All the high-throughput sequencing data have been deposited in Gene Expression Omnibus with the accession code GSE233781.

Acknowledgements

We thank the Arabidopsis Biological Resource Centre for providing the mutant seeds used in this study. The mutant seeds of *nrpd1-3*, *nrpe1-11*, and *drm1/2* were kindly provided by Dr. Shulin Deng. This work was supported by the National Natural Science Foundation of China to C.C. (No. 32070648) and J.S. (No. 32100474), Guangdong Pearl River Youth Talent Program to C.C. (No. 2021QN020018) and the Natural Science and Engineering Council of Canada to Y.C. (no. RGPIN/04625-2017).

Author contributions

C.C. conceived the project; C.C., C.H.W., and Y.J.L. designed the experiments; N.D. performed SPT6LdeltaWG/GW related ChIP-seq and immunoblots; J.S. performed the all the ChIP-qPCR assays; Z.Z. examined qPCR and Chop-PCR experiments; Y.J.L. performed all the rest of the experiments; Y.J.L., J.Y.L. and J.L. analyzed ChIP-seq data; C.C., and Y.H.C. analyzed rest of high-throughput data; C.C. and Y.J.L. wrote the paper.

Competing interests

The authors declare no competing interests.

References

1 Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483-508 (2013).
2 Obermeyer, S., Kapoor, H., Markusch, H. & Grasser, K. D. Transcript elongation by RNA polymerase II in plants: factors, regulation and impact on gene expression. *Plant J.*, 16115 (2023).
3 Sdano, M. A. et al. A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. *Elife* ***6*** (2017).
4 Aoi, Y. et al. SPT6 functions in transcriptional pause/release via PAF1C recruitment. *Mol. Cell* ***82***, 3412-3423 e3415 (2022).
5 Narain, A. et al. Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol. Cell* **81**, 3110-3127 (2021).
6 Zobeck, K. L., Buckley, M. S., Zipfel, W. R. & Lis, J. T. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* **40**, 965-975 (2010).
7 Dronamraju, R. et al. Spt6 Association with RNA Polymerase II Directs mRNA Turnover During Transcription. *Mol. Cell* **70**, 1054-1066 e1054 (2018).
8 Doris, S. M. et al. Spt6 Is Required for the Fidelity of Promoter Selection. *Mol. Cell* **72**, 687-699 e686 (2018).
9 Chen, C. et al. RNA polymerase II-independent recruitment of SPT6L at transcription start sites in Arabidopsis. *Nuclei Acids Res.* **47**, 6714-6725 (2019).
10 Gu, X. L., Wang, H., Huang, H. & Cui, X. F. SPT6L encoding a putative WG/GW-repeat protein regulates apical-basal polarity of embryo in Arabidopsis. *Mol. Plant* **5**, 249-259 (2012).
11 Shu, J., Ding, N., Liu, J., Cui, Y. & Chen, C. Transcription elongator SPT6L regulates the occupancies of the SWI2/SNF2 chromatin remodelers SYD/BRM and nucleosomes at transcription start sites in Arabidopsis. *Nuclei Acids Res.* **50**, 12754-12767 (2022).
12 Haag, J. R. & Pikaard, C. S. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**, 483-492 (2011).
13 Zhang, H., Lang, Z. & Zhu, J. K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489-506 (2018).
14 Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204-220 (2010).
15 Zhao, Y. & Chen, X. Noncoding RNAs and DNA Methylation in Plants. *Natl. Sci. Rev.* **1**, 219-229 (2014).
16 Zhai, J. et al. A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell* **163**, 445-455 (2015).

17 Huang, K. et al. Pol IV and RDR2: A two-RNA-polymerase machine that produces double-stranded RNA. *Science* **374**, 1579-1586 (2021).
18 Bohmdorfer, G. et al. Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* **5** e19092 (2016).
19 Haag, J. R. *et al.* In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. cell* **48** 811-818 (2012).
20 Xie, G. et al. Structure and mechanism of the plant RNA polymerase V. *Science* **379**, 1209-1213 (2023).
21 Antosz, W. et al. The Composition of the Arabidopsis RNA Polymerase II Transcript Elongation Complex Reveals the Interplay between Elongation and mRNA Processing Factors. *Plant Cell* **29**, 854-870 (2017).
22 Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612 (2018).
23 Gallego-Bartolome, J. et al. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. *Cell* **176**, 1068-1082 e1019 (2019).
24 Ream, T. S. *et al.* Subunit compositions of Arabidopsis RNA polymerases I and III reveal Pol I- and Pol III-specific forms of the AC40 subunit and alternative forms of the C53 subunit. *Nuclei Acids Res.* **43**, 4163-4178 (2015).
25 He, L. et al. DNA methylation-free Arabidopsis reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat. Commun.* **13**, 1335 (2022).
26 Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* **152**, 352-364 (2013).
27 Zheng, B. et al. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev.* **23**, 2850-2860 (2009).
28 Law, J. A. et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385-389 (2013).
29 Lahmy, S. et al. Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev.* **30**, 2565-2570 (2016).
30 Karlowski, W. M. et al. Genome-wide computational identification of WG/GW Argonaute-binding proteins in Arabidopsis. *Nuclei Acids Res.* **38**, 4231-4245 (2010).
31 Tsuzuki, M. et al. Broad noncoding transcription suggests genome surveillance by RNA polymerase V. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 30799-30804 (2020).
32 Liu, W. et al. RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis. *Nat. Plants* **4**, 181-188 (2018).
33 Rowley, M. J., Avrutsky, M. I., Sifuentes, C. J., Pereira, L. & Wierzbicki, A. T. Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. *PLoS Genet.* **7**, e1002120 (2011).
34 Onodera, Y. et al. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613-622 (2005).
35 Zhong, X. et al. Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 911-916 (2015).
36 Gao, Z. et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106-109 (2010).
37 He, X. J. et al. A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. *Genes Dev.* **23**, 2717-2722 (2009).

38 Zhong, X. et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870-875 (2012).
39 Erdmann, R. M. & Picard, C. L. RNA-directed DNA Methylation. *PLoS Genet.* **16**, e1009034 (2020).
40 Kollen, K. et al. The zinc-finger protein SPT4 interacts with SPT5L/KTF1 and modulates transcriptional silencing in Arabidopsis. *FEBS Lett.* **589**, 3254-3257 (2015).
41 Wulf, M. G. et al. Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *J. Biol. Chem.* **294**, 18220-18231 (2019).
42 Wendte, J. M. & Pikaard, C. S. The RNAs of RNA-directed DNA methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860,** 140-148 (2017).
43 He, X. J. et al. An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**, 498-508 (2009).
44 Pontier, D. et al. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**, 2030-2040 (2005).
45 Henderson, I. R. & Jacobsen, S. E. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* **22,** 1597-1606 (2008).
46 Bies-Etheve, N. et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**, 649-654 (2009).
47 Haag, J. R., Pontes, O. & Pikaard, C. S. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One* **4**, e4110 (2009).
48 Wendte, J. M. et al. Functional Dissection of the Pol V Largest Subunit CTD in RNA-Directed DNA Methylation. *Cell Rep.* **19**, 2796-2808 (2017).
49 Chen, C. et al. Cytosolic acetyl-CoA promotes histone acetylation predominantly at H3K27 in Arabidopsis. *Nat. Plants* **3**, 814-824 (2017).
50 Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, e46314 (2019).
51 Rowley, M. J., Bohmdorfer, G. & Wierzbicki, A. T. Analysis of long non-coding RNAs produced by a specialized RNA polymerase in Arabidopsis thaliana. *Methods* **63**, 160-169 (2013).
52 Schon, M. A., Kellner, M. J., Plotnikova, A., Hofmann, F. & Nodine, M. D. NanoPARE: parallel analysis of RNA 5' ends from low-input RNA. *Genome Res.* **28**, 1931-1942 (2018).
53 Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J. Comput. Biol.* **24**, 1138-1143 (2017).
54 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.* Methods **9**, 357-359 (2012).
55 Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
56 Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728-1740 (2012).
57 Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nuclei Acids Res.* **44**, W160-165 (2016).
58 Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478-2492 (2017).
59 Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422-423 (2021).

60 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).

61 Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).

62 Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491-499 (2017).

Figure1. SPT6L co-occupies and interacts with NRPE1 in Arabidopsis

A: Genome tracks display SPT6L, Pol II, and NRPE1 ChIP-seq signals on chromosome 1 (Chr1: 6,700 to 6,780 kb). The ChIP signal of each sample was averaged over two biological replicates. The y-axis values indicate the mean of normalized reads per 10 bp. The black arrows pointed to co-binding peaks between SPT6L and NRPE1.

B: Heatmaps of SPT6L, Pol II, and NRPE1 ChIP signals around peak center of all SPT6L peaks. The SPT6L peaks were clustered into two groups (genic and intergenic). The plotted regions are upstream and downstream 1 kb of peak center.

C: Binding profiles of SPT6L, Pol II, and NRPE1 at characterized six genomic groups. The six groups of regions were clustered according to the solo/double enrichment among the three proteins (see Supplementary Figure 2D). The ChIP signal of each sample was averaged over two biological replicates. The plotted regions were 2 kb around the center of regions (upstream and downstream 1 kb, respectively). The y-axis value indicates the relative mean of normalized reads (1× sequencing depth normalization) per 10 bp non-overlapping bins. The number of regions in each group (G1 to G6) were indicated in the graph.

D: Heatmaps of the distance between enriched regions in each state and transcription start sites (TSS). Each rectangle represents 200 bp.

E: Co-immunoprecipitation (Co-IP) examined the interaction between SPT6L and core subunits of Pol V complex. Immunoprecipitation (IP) and Western blot (WB) were performed using specified antibodies. Data from two biological replicates were shown.

Figure 2. NRPE1 is required for the intergenic enrichment of SPT6L

A: Heatmaps of NRPE1 ChIP signals in *nrpe1 NRPE1-GFP* (*NRPE1*) and *nrpe1 spt6l NRPE1-GFP* (*NRPE1 spt6l*) background. The plotted regions were similar to Figure 1C and the strength of ChIP signals at 1 kb around of region centers were shown.

B: Scatterplot of NRPE1 ChIP signals in *NRPE1* and *NRPE1 spt6l* at Pol V peaks. ChIP signals ($\log_2$ values) in *NRPE1* (y-axis) and *NRPE1 spt6l* (x-axis) were plotted.

C: Binding profiles of SPT6L in WT and *nrpe1* at four previously defined genomic groups. The ChIP signal of each sample was averaged over two biological replicates. The regions were plotted as indicated in Figure 1C.

D: Binding profiles of SPT6L occupancy in WT and nrpe1 at the nearest downstream TSS of four genomic states. Plotting regions were scaled to the same length as follows: 5' ends (–1.0 kb to TSS) and 3' ends (transcription termination site [TTS] to downstream 1.0 kb) were not scaled, and the gene bodies were scaled to 2 kb. The y-axis was plotted as described in Figure 1C. The number of genes were indicated (n).

E: Binding profiles of SPT6L occupancy at the four genomic states after 1 h mock and Flavopiridol (FLA) treatment. The regions were plotted as indicated in Figure 1C.

Figure 3. SPT6L is involved in the regulation of DNA methylation

A: Chop-PCR analysis of DNA methylation at *SN1*, *IGN5*, *IGN23*, and *IGN25* performed by digestion with *Hae*III restriction endonuclease. Digested genomic DNA was amplified by PCR. Sequences lacking *Hae*III (*Actin2*) were used as loading controls.

B: Boxplots of DNA methylation level at three contexts (CG, CHG, and CHH) among *WT*, *nrpe1*, *spt6l*, and *nrpe1 spt6l*. The plotted regions were *spt6l* DMRs and Data from two biological replicates were shown.

C: Genome tracks display the ChIP-seq and BS-seq signals on chromosome 1 (Chr1: 6,700 to 6,780 kb). The ChIP signals included SPT6L and NRPE1 in WT/*nrpe1* and WT/*spt6l* background, respectively. The BS-seq signals showed total methylation levels in *WT*, *nrpe1*, *spt6l*, and *nrpe1 spt6l*. Each sample was averaged over two biological replicates. The red rectangle highlighted SPT6L and NRPE1 co-targeted loci with changed DNA methylation levels in different genetic background.

D: Boxplots of DNA methylation level at three contexts (CG, CHG, and CHH) among *WT*, *nrpe1*, *spt6l*, and *nrpe1 spt6l*. The plotted regions were previously defined six genomic groups and Data from two biological replicates were shown.

E: Heatmap showed the expression of RdDM pathway related genes by RNA-seq. The threshold to define differentially expressed genes is more than 2-fold expression change (*F*) and adjust *p*-value (*P*)less than 0.01. The genes with either less than 2-fold change or p-value larger than 0.01 were labeled as gray. Three biological replicates were included.

F: Stacked bar graph showed the proportion of different size of small RNA in WT and *spt6l*. Data from three biological replicates were shown.

G: Boxplots showed the amount of 24-nt siRNA in WT and *spt6l* at Pol IV and V dependent regions. The amount of 24-nt siRNA (y-axis) Stacked bar graph showed the proportion of different size of small RNA in WT and *spt6l*. Data from three biological replicates were shown.

Figure 4. The WG/GW repeat of SPT6L is dispensable for its intergenic enrichment and DNA methylation

A: The morphological phenotypes of 7-day old *WT*, *spt6l*, *spt6l pSPT6L:SPT6L-GFP*, and *spt6l pSPT6L:SPT6LΔWG/GW* seedlings. Bar = 0.5 mm

B: Heatmaps of SPT6L and SPT6LΔWG/GW ChIP signals at SPT6L binding peaks. The plotted regions were similar to Figure 1B and the strength of ChIP signals at 1 kb around of region centers were shown.

C: Scatterplot of SPT6L and SPT6LΔWG/GW ChIP signals in genomic binds. The Arabidopsis genome were divided into 100 bp length bins. ChIP signals (log_2 values) in SPT6LΔWG/GW (y-axis) and SPT6L (x-axis) were plotted.

D: Genome tracks display the DNA methylation signals (in CG, CHG, and CHH contexts) of *WT* and *spt6l SPT6LΔWG/GW* on part of chromosome 5. Data from one biological replicate was shown.

E: Plots of DNA methylation levels in *WT* and *spt6l SPT6LΔWG/GW* within the genic and intergenic peaks of SPT6L. The plots displayed DNA methylation in CG, CHG, and CHH context. Data from one (*spt6l SPT6LΔWG/GW*) or two (*WT*) biological replicate(s) were shown.

Figure 5. SPT5L is required for the intergenic enrichment of SPT6L

A: Genome tracks display the ChIP-seq signals of SPT6L, SPT6L *spt5l*, and NRPE1 on chromosome 1 (Chr1: 6,700 to 6,780 kb). Each sample was averaged over two biological replicates.

B: Binding profiles of SPT6L in *WT* and *spt5l* at four previously defined genomic groups. The ChIP signal of each sample was averaged over two biological replicates. The regions were plotted as indicated in Figure 1C.

C: ChIP–qPCR showing genomic occupancy by NRPE1-GFP fusion protein in *NRPE1-GFP* and *spt5l NRPE1-GFP*. All the fold changes are relative to ChIP signal obtained at *ACT7* in each sample and replicates. Error bars are presented as mean values ± s.d. from three biological replicates. All significant differences were indicated with $*P < 0.05$, $**P < 0.01$ (unpaired, two-tailed Student's *t*-test).

D: Co-IP examined the interaction between SPT6L and SPT5L. Immunoprecipitation (IP) and Western blot (WB) were performed using specified antibodies. Data from two biological replicates were shown.

E: Co-IP examined the role of SPT5L in the interaction between SPT6L and NRPE1. Immunoprecipitation (IP) and Western blot (WB) were performed using specified antibodies. Data from two biological replicates were shown.

F: ChIP–qPCR showing genomic occupancy by SPT6L-GFP fusion protein in *SPT6L-GFP*, *drm1 drm2 SPT6L-GFP*, and *nrpd1 SPT6L-GFP*. All the fold changes are relative to ChIP signal obtained at *ACT7* in each sample and replicates. Error bars are presented as mean values ± s.d. from three biological replicates. Different lowercase letters indicate significant differences, as determined by one-way *ANOVA*, $P < 0.05$.

Figure 6. SPT6L is required to maintain Pol V transcripts

A: Genome tracks display the ChIP-seq and RIP-seq signals on chromosome 5. The RIP-seq signals of each sample from different strands were displayed as positive and negative values. The signals from *nrpe1* RIP-seq sample were serves as background. The black arrows pointed to the loci with changed NRPE1 RIP signals in *spt6l*. The ChIP-seq or RIP-seq signals of each sample were averaged over two biological replicates.

B: Boxplots showed the amount of Pol V transcripts in *NRPE1-GFP* and *spt6l NRPE1-GFP* within Pol V peaks. The reads number were normalized to total mapped RIP-seq reads of each sample. Unpaired Wilcoxon test was performed. $P < 2.2e^{-16}$. Two biological replicates were included.

C: Profiles of normalized NRPE1 RIP-seq reads from *NRPE1-GFP* and *spt6l NRPE1-GFP* within four previous defined genomic groups (Figure 1C).

D: Boxplots showed the reads length of NRPE1 RIP-seq from *NRPE1-GFP* and *spt6l NRPE1-GFP*. Unpaired Wilcoxon test was performed. $P < 2.2e^{-16}$. Two biological replicates were included.

E: The relative nucleotide bias of each position in the upstream and downstream 20-nt of nascent transcripts captured in *NRPE1-GFP* and *spt6l NRPE1-GFP*.

Supplementary Figure 1. SPT6L co-binds with NRPE1 at intergenic regions

A: Heatmaps of SPT6L, Pol II, and NRPE1 ChIP signals around peak center of all SPT6L peaks. The SPT6L peaks were clustered into two groups (genic and intergenic). The plotted regions are upstream and downstream 1 kb of peak center. The SPT6L and Pol II ChIP-seq data sourced from GSE108673. The NRPE1 ChIP-seq data sourced from GSE124546.

B: Pie charts showed the proportions of transposable elements (TE), gene, and others within SPT6L genic and intergenic peaks.

C: Heatmaps of NRPE1 and SPT6L ChIP signals around peak center of all NRPE1 peaks. According to the differential binding patterns of SPT6L at NRPE1 peaks, the NRPE1 peaks had been clustered into NRPE1-only and NRPE1-SPT6L overlapped peaks.

D: The plots of TE frequency within NRPE1-SPT6L overlapped and random peaks (defined in Supplementary Figure 1C). Two types of randomizations were applied: randomly selected same amounts of peaks as that of NRPE1-SPT6L overlapped peaks in total NRPE1 binding sites (blue lines); randomly shuffled NRPE1-SPT6L overlapped peaks across entire genome (gray lines).

E: Pie charts indicated the proportions of different TE groups in Arabidopsis genome, NRPE1 binding peaks, and NRPE1-SPT6L overlapped peaks, respectively.

Supplementary Figure 2. HMM model defined genome states and the interaction of Pol V and SPT6L

A: Confocal image examined the nuclear localization of NRPE1–GFP in 7-day after germination (DAG) roots; Scale bars, 20 μm.

B: Chop-PCR analysis of DNA methylation at *SN1*, *IGN23*, and *IGN25* in *WT*, *nrpe1*, and three transgenic lines of *nrpe1 ProNRPE1:NRPE1-GFP* performed by digestion with *Hae*III restriction endonuclease. Digested genomic DNA was amplified by PCR. Sequences lacking *Hae*III (*Actin2*) were used as loading controls.

C: Venn diagram showed the reproducibility of our identified NRPE1 binding peaks (in seedlings) with published NRPE1 peak list (inflorescence).

D: Heatmap showed six genomic states defined by ChromHMM according to differential enrichment of SPT6L, Pol II, and NRPE1.

E: Heatmap showed the enrichment of different genomic features within six defined genomic states.

F: Yesat-two hybrid assay to examine the direct interaction between SPT6L and multiple Pol V subunits. Growth of transformed yeast is shown on permissive $SD^{-His-Leu-Trp}$ plus 3AT medium. For each pair of AD and BD constructs, three different dilutions (x1, x10, and x100) were shown.

G: Confocal images examined the nuclear localization of NRPE4-GFP and NRPE7-GFP in 7 DAG roots; Scale bars, 20 μm.

Supplementary Figure 3. Pol V is required for the intergenic enrichment of SPT6L

A: ChIP–qPCR showing genomic occupancy by NRPE1-GFP fusion protein in *NRPE1-GFP* and *spt6l NRPE1-GFP*. All the fold changes are relative to ChIP signal obtained at *ACT7* in each sample and replicates. Error bars are presented as mean values ± s.d. from

three biological replicates. All significant differences were indicated with *P < 0.05, **P < 0.01 (unpaired, two-tailed Student's *t*-test).

B: Immunoblot assessing the levels of NRPE1 proteins in the genetic backgrounds as indicated. H3 levels served as loading controls. Data from two biological replicates were shown.

C: ChIP–qPCR showing genomic occupancy by SPT6L-GFP fusion protein in *SPT6L-GFP* and *nrpe1 SPT6L-GFP*. All the fold changes are relative to ChIP signal obtained at *ACT7* in each sample and replicates. Error bars are presented as mean values ± s.d. from three biological replicates. All significant differences were indicated with *P < 0.05, **P < 0.01 (unpaired, two-tailed Student's *t*-test).

D: Immunoblot assessing the levels of SPT6L proteins in the genetic backgrounds as indicated. H3 levels served as loading controls. Data from two biological replicates were shown.

Supplementary Figure 4. SPT6L is involved in the regulation of DNA methylation

A: Volcano plot highlighted the differentially expressed genes (DGEs)between WT and *spt6l*. The increased and decreased genes in *spt6l* were labelled with red and blue, respectively. Grey dots represented the genes were failed to pass the threshold (|Fold Change| ≥ 2 and adjust *p*-value < 0.01. Three biological replicates were included.

B: Heatmap of methylation levels of indicated mutants within *spt6l* CG, CHG, and CHH hypomethylation DMRs. Genotypes (columns) have also been clustered. The plotted values were the relative methylation level in mutants to that in WT.

Supplementary Figure 5. Stable transgenic line of *spt6l SPT6LΔWG/GW-GFP*

A: Confocal image examined the localization of SPT6L-GFP and SPT6LΔWG/GW-GFP in 7-day after germination (DAG) roots; Scale bars, 20 μm.

B: Immunoblot assessing the levels of SPT6L and SPT6LΔWG/GW proteins. The image of Coomassie blue staining serves as loading control.

Supplementary Figure 6. Protein amounts and the interaction between SPT6L and SPT5L

A: ChIP–qPCR showing genomic occupancy by SPT6L-GFP fusion protein in *SPT6L-GFP* and *spt5l SPT6L-GFP*. All the fold changes are relative to ChIP signal obtained at *ACT7* in each sample and replicates. Error bars are presented as mean values ± s.d. from three biological replicates. All significant differences were indicated with *P < 0.05, **P < 0.01 (unpaired, two-tailed Student's *t*-test).

B: Immunoblot assessing the levels of SPT6L proteins in the *WT* and *spt5l* as indicated. H3 levels served as loading controls. Data from three biological replicates were shown.

C: Yesat-two hybrid assay to examine the direct interaction between SPT5L and SPT6L. The diagram on the left showed the different domains of SPT5L and designed truncations of SPT5L in following Yeast-two hybrid assays. Growth of transformed yeast is shown on permissive $SD^{-His-Leu-Trp}$ plus 3AT medium. For each pair of AD and BD constructs, three different dilutions (x1, x10, and x100) were shown.

D: Immunoblot assessing the levels of SPT6L proteins in the *WT*, *drm1 drm2*, and *nrpd1* as indicated. H3 levels served as loading controls. Data from three biological replicates were shown.

Supplementary Figure 7. Pearson correlations of replicates in ChIP-seq and smRNA-seq data.

A: The correlation values were calculated based on reads numbers of different samples within all SPT6L or NRPE1 peaks.

B: The entire genome was divided into equal bins (100 bp in length) and the numbers of RNA-seq and 24-nt smRNA-seq reads from *WT* and *spt6l* were calculated in each bin. The correlation values were calculated based on reads numbers in each bin between *WT* and *spt6l*.

Supplementary Figure 8. Pearson correlations of replicates in BS-seq data.

The entire genome was divided into equal bins (500 bp in length) and the methylation levels in three different contexts from *WT*, *nrpe1*, *spt6l*, and *nrpe1 spt6l* were calculated in each bin.

Figure 1

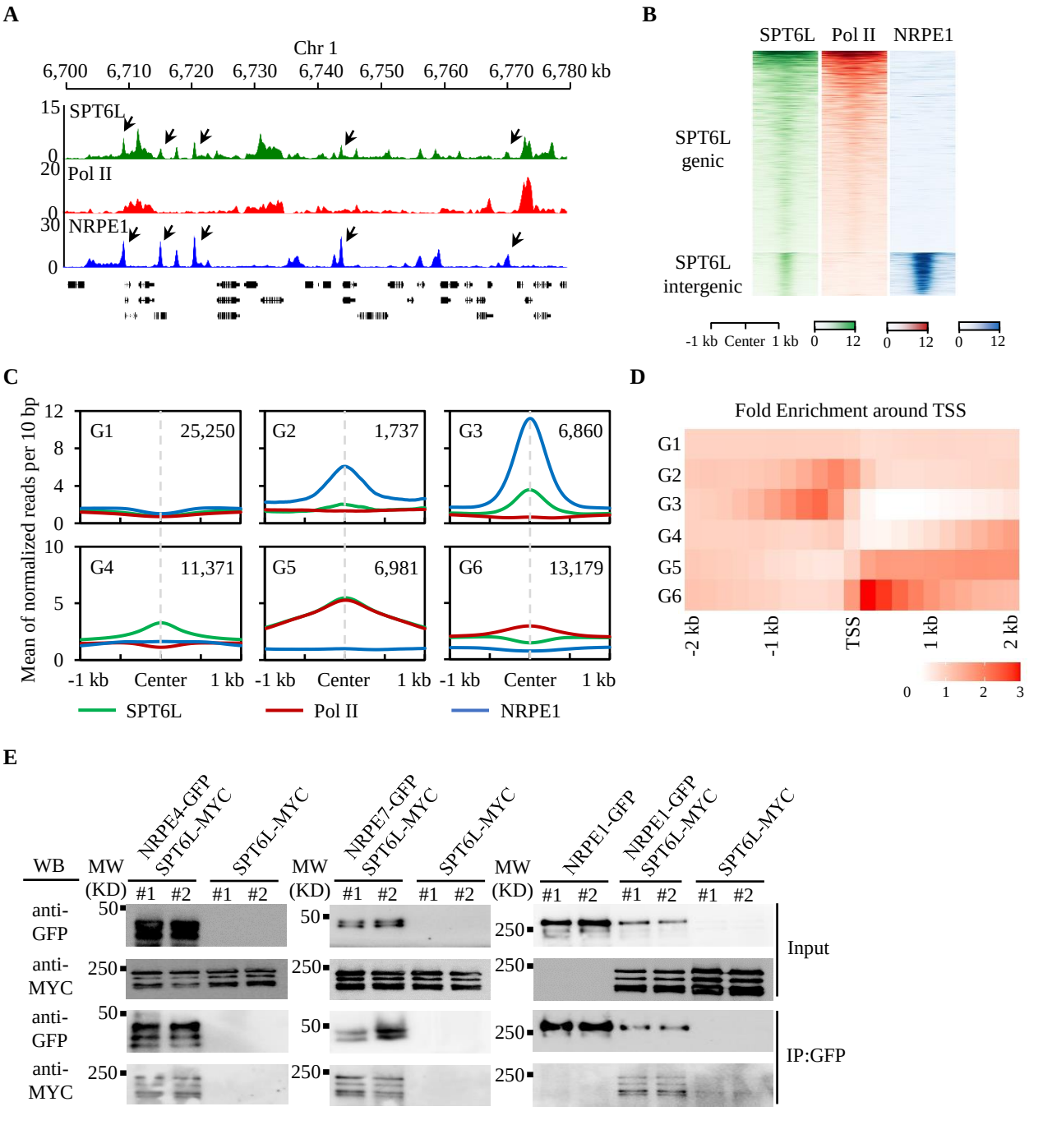

A
Chr 1
6,700 6,710 6,720 6,730 6,740 6,750 6,760 6,770 6,780 kb
SPT6L
Pol II
NRPE1
B
SPT6L Pol II NRPE1
SPT6L genic
SPT6L intergenic
-1 kb Center 1 kb
C
Mean of normalized reads per 10 bp
G1 25,250
G2 1,737
G3 6,860
G4 11,371
G5 6,981
G6 13,179
-1 kb Center 1 kb
SPT6L
Pol II
NRPE1
D
Fold Enrichment around TSS
G1
G2
G3
G4
G5
G6
-2 kb
-1 kb
TSS
1 kb
2 kb
E
WB
MW (KD)
NRPE4-GFP SPT6L-MYC
SPT6L-MYC
NRPE7-GFP SPT6L-MYC
NRPE1-GFP
NRPE1-GFP SPT6L-MYC
#1 #2
anti-GFP
anti-MYC
Input
IP:GFP

Figure 2

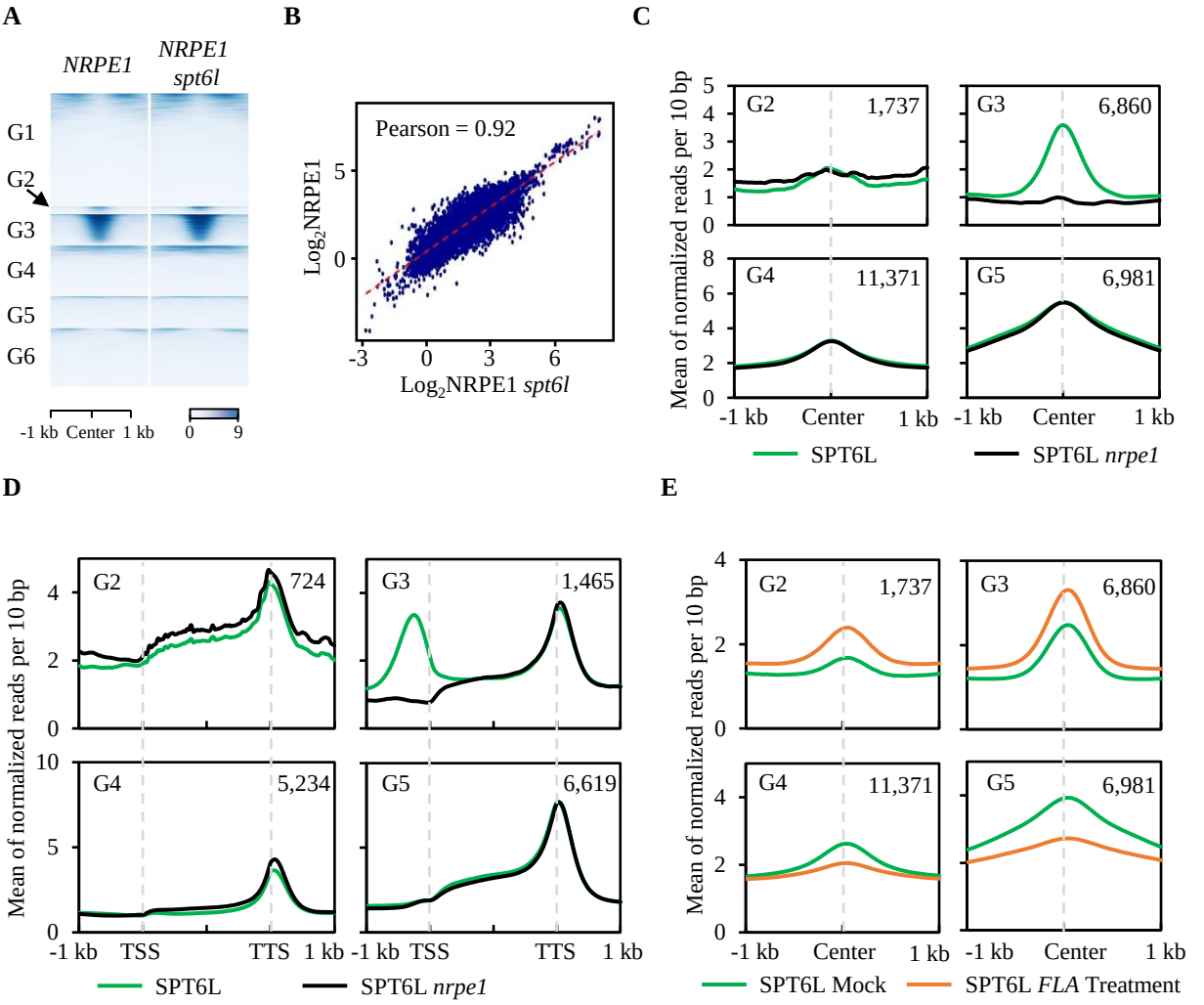

A
NRPE1
NRPE1 spt6l
G1
G2
G3
G4
G5
G6
-1 kb Center 1 kb
0 9
B
Pearson = 0.92
Log2NRPE1
Log2NRPE1 spt6l
C
Mean of normalized reads per 10 bp
G2 1,737
G3 6,860
G4 11,371
G5 6,981
-1 kb Center 1 kb
SPT6L
SPT6L nrpe1
D
G2 724
G3 1,465
G4 5,234
G5 6,619
-1 kb TSS TTS 1 kb
SPT6L
SPT6L nrpe1
E
G2 1,737
G3 6,860
G4 11,371
G5 6,981
-1 kb Center 1 kb
SPT6L Mock
SPT6L FLA Treatment

Figure 3

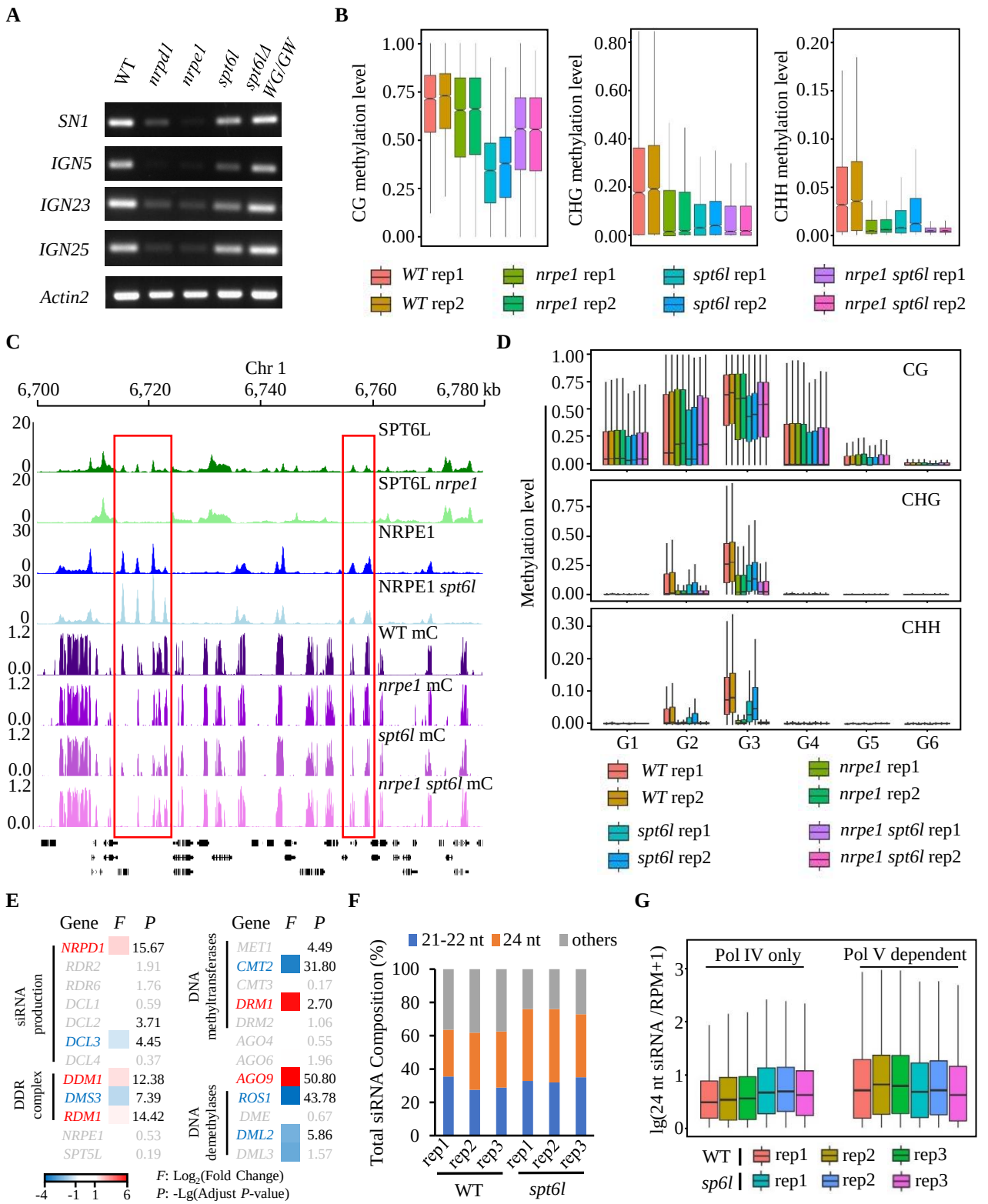

A
WT
nrpd1
nrpe1
spt6l
spt6lΔ WG/GW
SN1
IGN5
IGN23
IGN25
Actin2
B
CG methylation level
CHG methylation level
CHH methylation level
WT rep1
WT rep2
nrpe1 rep1
nrpe1 rep2
spt6l rep1
spt6l rep2
nrpe1 spt6l rep1
nrpe1 spt6l rep2
C
Chr 1
6,700
6,720
6,740
6,760
6,780 kb
SPT6L
SPT6L nrpe1
NRPE1
NRPE1 spt6l
WT mC
nrpe1 mC
spt6l mC
nrpe1 spt6l mC
D
Methylation level
CG
CHG
CHH
G1
G2
G3
G4
G5
G6
E
Gene F P
siRNA production
NRPD1 15.67
RDR2 1.91
RDR6 1.76
DCL1 0.59
DCL2 3.71
DCL3 4.45
DCL4 0.37
DDR complex
DDM1 12.38
DMS3 7.39
RDM1 14.42
NRPE1 0.53
SPT5L 0.19
DNA methyltransferases
MET1 4.49
CMT2 31.80
CMT3 0.17
DRM1 2.70
DRM2 1.06
AGO4 0.55
AGO6 1.96
AGO9 50.80
DNA demethylases
ROS1 43.78
DME 0.67
DML2 5.86
DML3 1.57
F: Log2(Fold Change)
P: -Lg(Adjust P-value)
F
21-22 nt
24 nt
others
Total siRNA Composition (%)
rep1
rep2
rep3
WT
spt6l
G
Pol IV only
Pol V dependent
lg(24 nt siRNA /RPM+1)
WT rep1 rep2 rep3
sp6l rep1 rep2 rep3

Figure 4

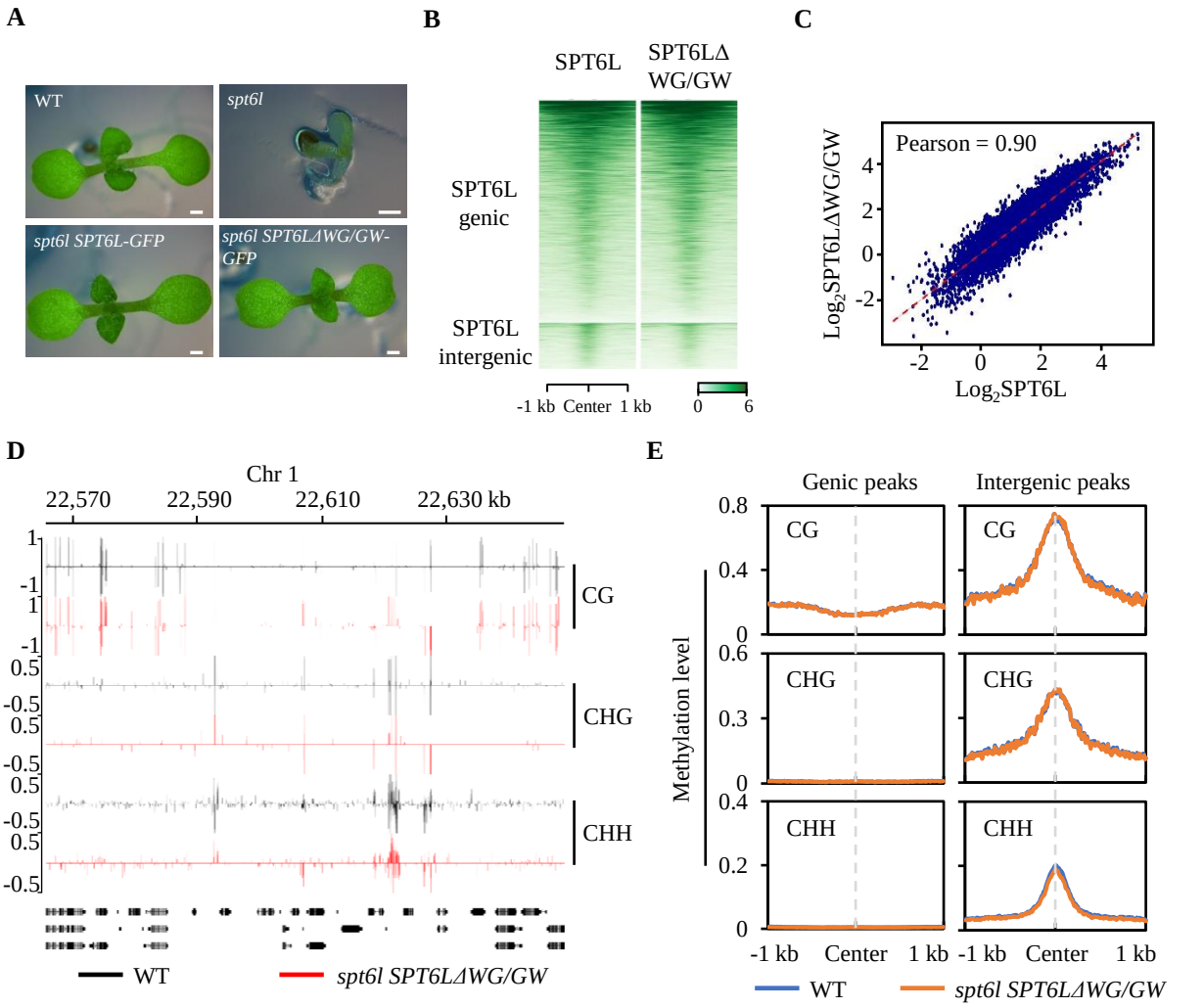

A
WT
spt6l
spt6l SPT6L-GFP
spt6l SPT6LΔWG/GW-GFP
B
SPT6L
SPT6LΔ WG/GW
SPT6L genic
SPT6L intergenic
-1 kb Center 1 kb
0 6
C
Pearson = 0.90
Log2SPT6LΔWG/GW
Log2SPT6L
D
Chr 1
22,570 22,590 22,610 22,630 kb
CG
CHG
CHH
WT
spt6l SPT6LΔWG/GW
E
Genic peaks
Intergenic peaks
Methylation level
CG
CHG
CHH
-1 kb Center 1 kb
WT
spt6l SPT6LΔWG/GW

Figure 5

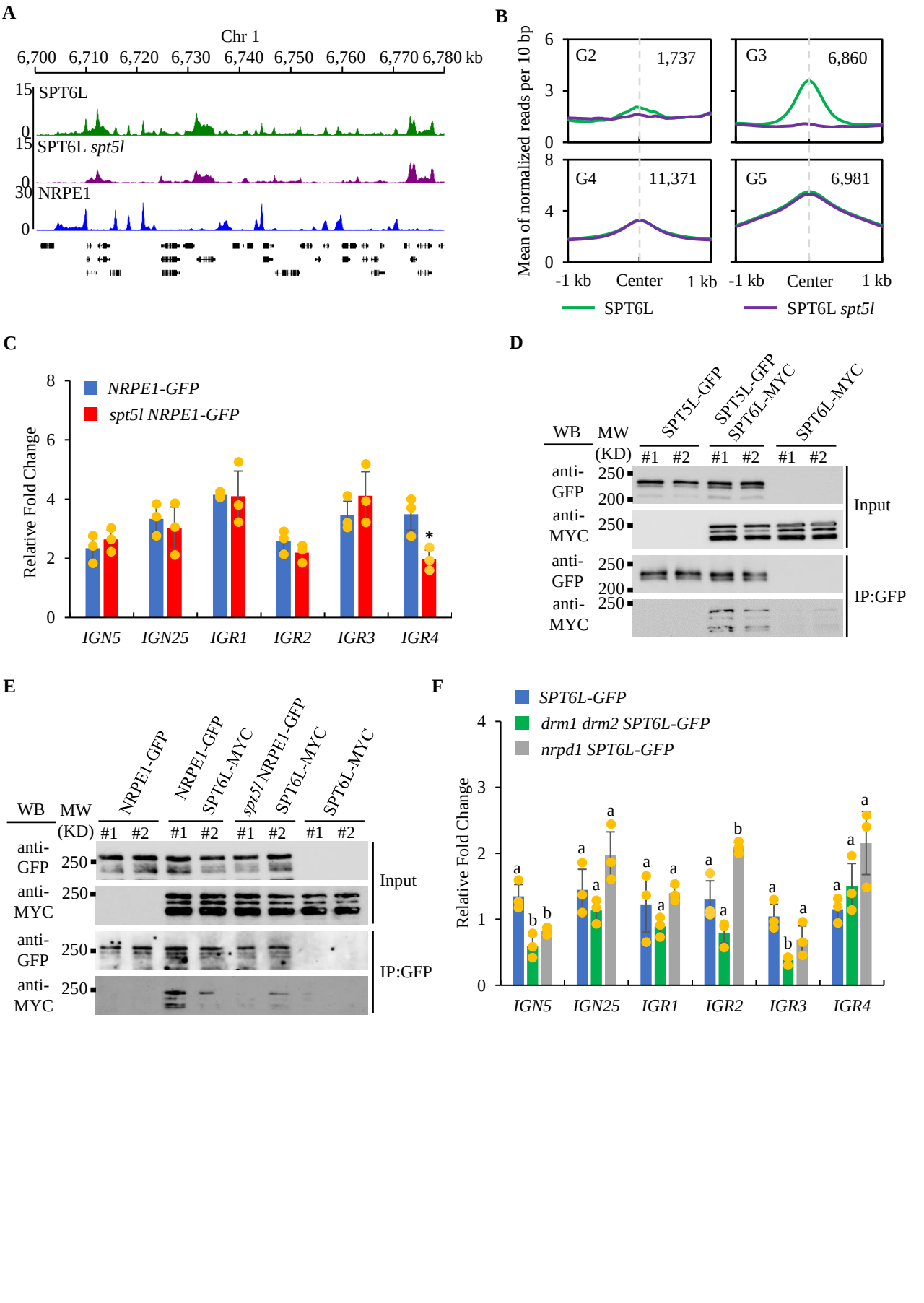
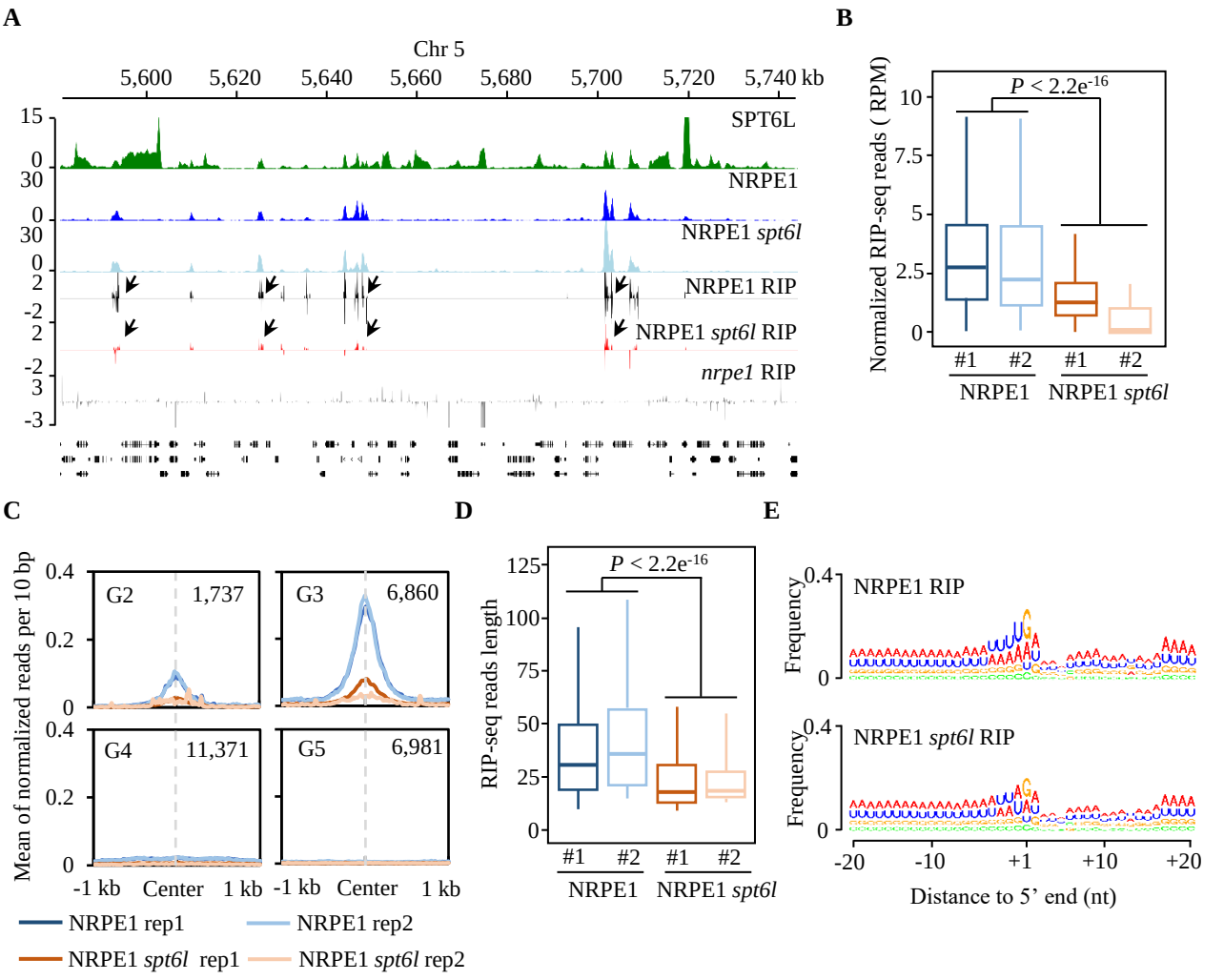
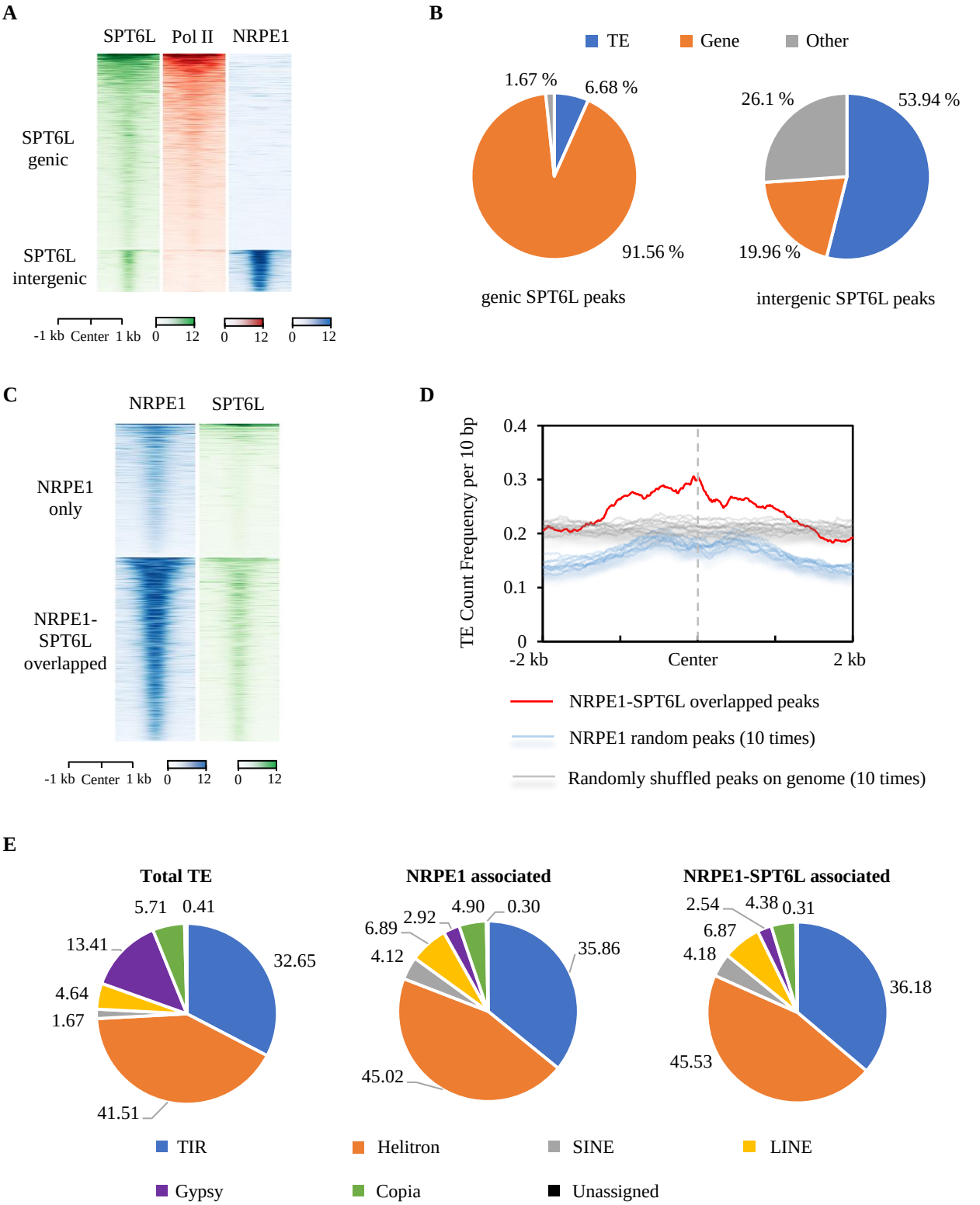

A
Chr 1
6,700 6,710 6,720 6,730 6,740 6,750 6,760 6,770 6,780 kb
SPT6L
SPT6L spt5l
NRPE1
B
Mean of normalized reads per 10 bp
G2 1,737
G3 6,860
G4 11,371
G5 6,981
-1 kb Center 1 kb
SPT6L
SPT6L spt5l
C
NRPE1-GFP
spt5l NRPE1-GFP
Relative Fold Change
IGN5 IGN25 IGR1 IGR2 IGR3 IGR4
D
SPT5L-GFP
SPT5L-GFP SPT6L-MYC
SPT6L-MYC
WB
MW (KD)
anti-GFP
anti-MYC
Input
IP:GFP
E
NRPE1-GFP
NRPE1-GFP SPT6L-MYC
spt5l NRPE1-GFP SPT6L-MYC
SPT6L-MYC
F
SPT6L-GFP
drm1 drm2 SPT6L-GFP
nrpd1 SPT6L-GFP
Relative Fold Change
IGN5 IGN25 IGR1 IGR2 IGR3 IGR4

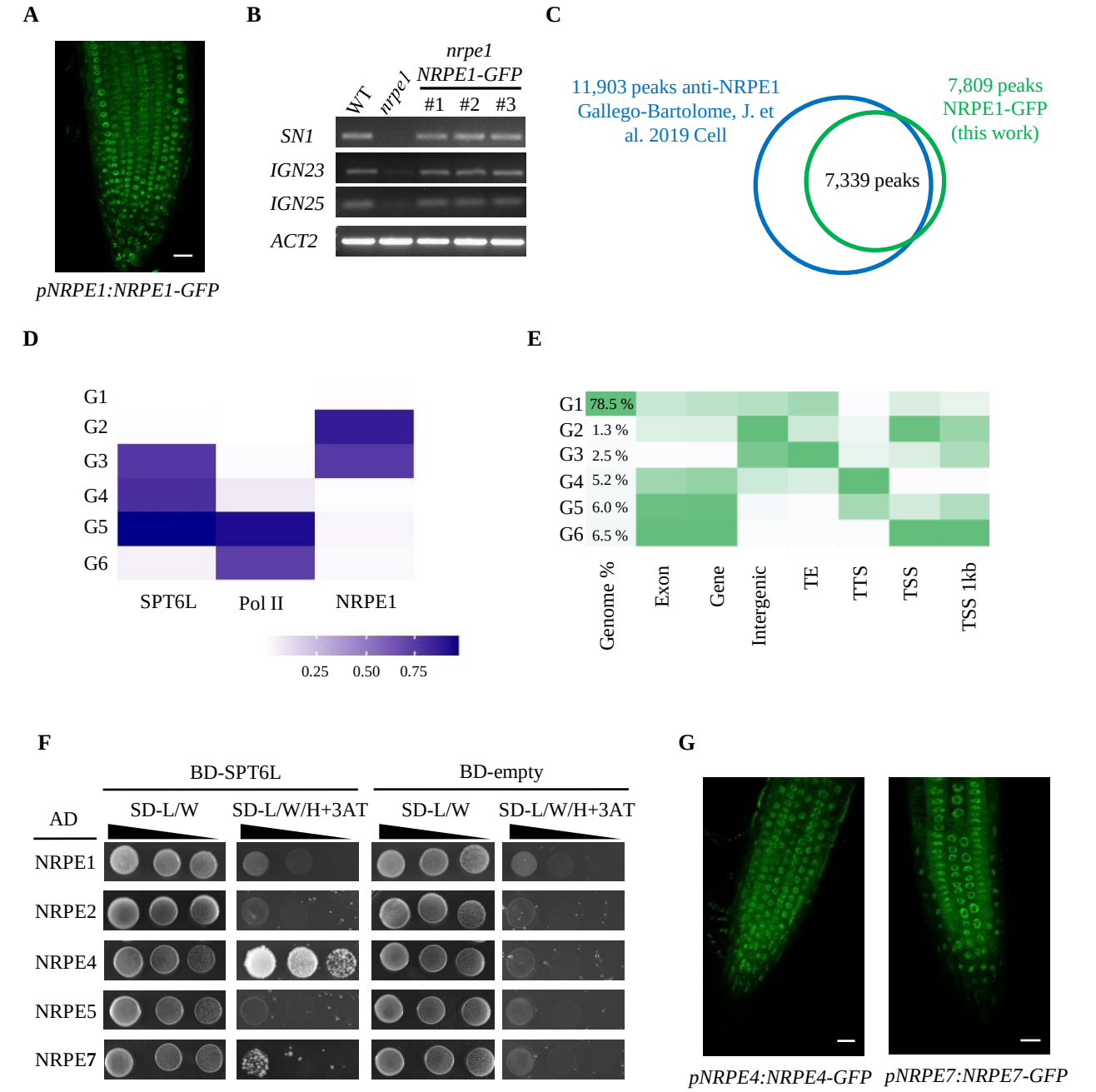
Figure 6


A
Chr 5
5,600
5,620
5,640
5,660
5,680
5,700
5,720
5,740 kb
SPT6L
NRPE1
NRPE1 spt6l
NRPE1 RIP
NRPE1 spt6l RIP
nrpe1 RIP
B
P < 2.2e-16
Normalized RIP-seq reads (RPM)
#1
#2
#1
#2
NRPE1
NRPE1 spt6l
C
Mean of normalized reads per 10 bp
G2
1,737
G3
6,860
G4
11,371
G5
6,981
-1 kb
Center
1 kb
NRPE1 rep1
NRPE1 rep2
NRPE1 spt6l rep1
NRPE1 spt6l rep2
D
P < 2.2e-16
RIP-seq reads length
E
NRPE1 RIP
NRPE1 spt6l RIP
Frequency
Distance to 5' end (nt)

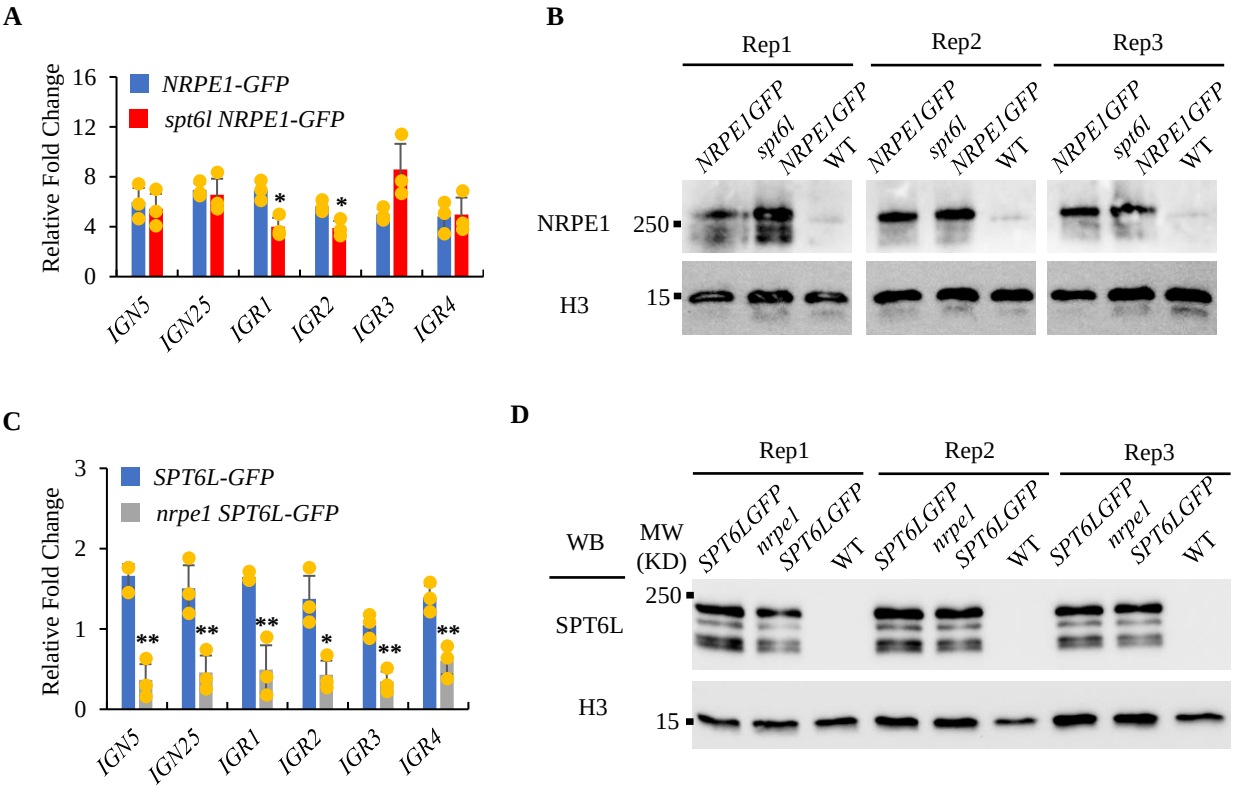
Supplementary Figure 1



Supplementary Figure 2

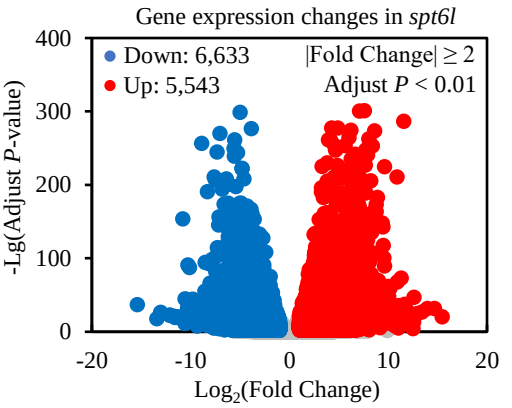


Supplementary Figure 3

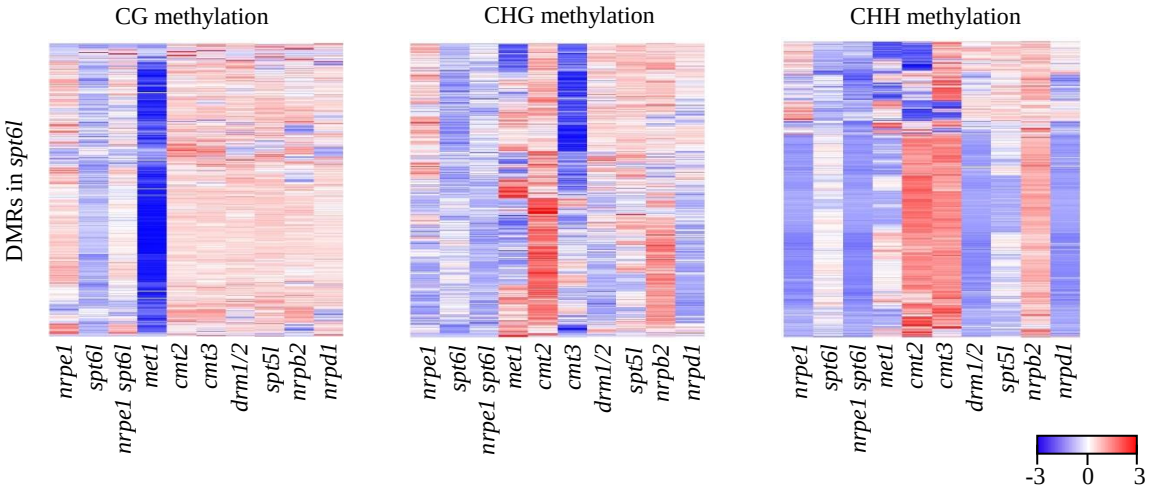


Supplementary Figure 4

A

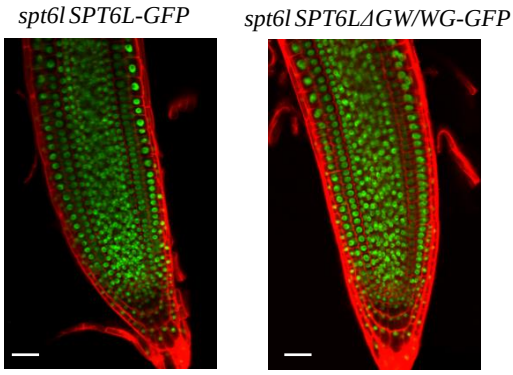


B

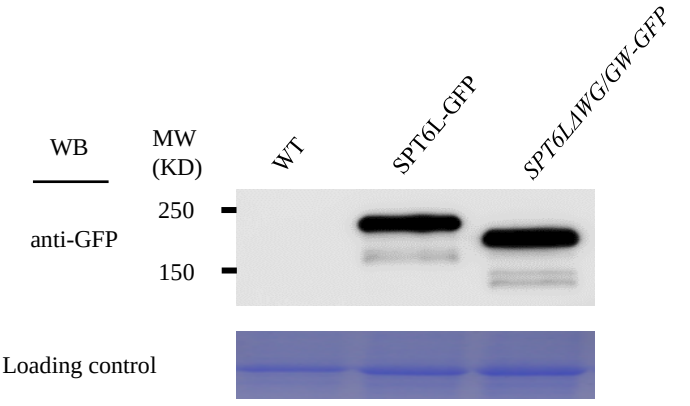


Supplementary Figure 5

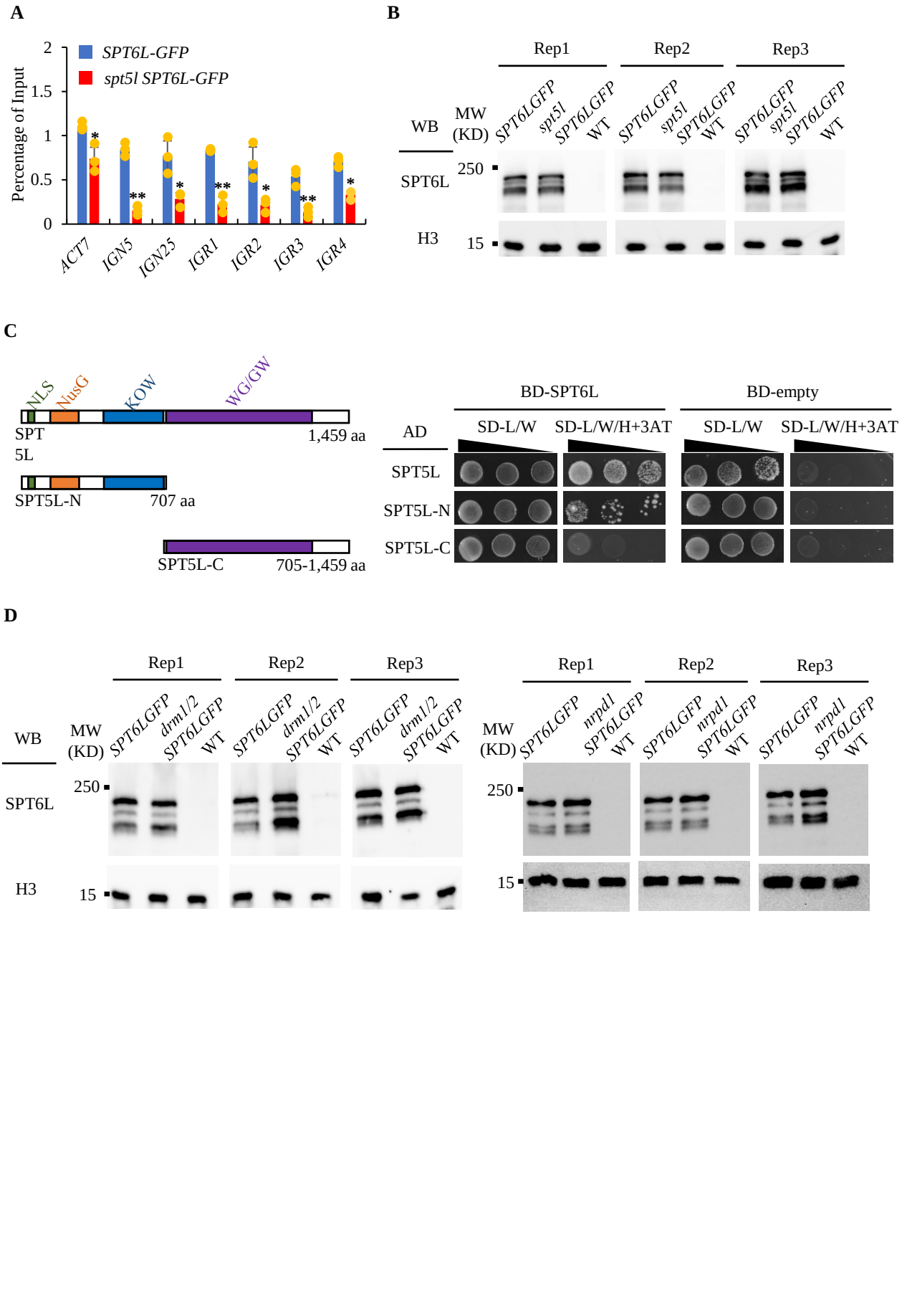
A



B



Supplementary Figure 6



Supplementary Figure 7

A

	Rep1	Rep2
Rep1	1.00	0.90
Rep2	0.90	1.00

nrpe1 NRPE1 GFP IP

	Rep1	Rep2
Rep1	1.00	0.99
Rep2	0.99	1.00

spt6l NRPE1 GFP IP

	Rep1	Rep2
Rep1	1.00	0.73
Rep2	0.73	1.00

nrpe1 SPT6L GFP IP

	Rep1	Rep2
Rep1	1.00	0.75
Rep2	0.75	1.00

spt51 SPT6L GFP IP

	Rep1	Rep2
Rep1	1.00	0.96
Rep2	0.96	1.00

SPT6LdeltaWG/GW GFP IP

B

		WT			*spt6l*		
		Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
WT	Rep1	1.00	1.00	0.99	0.56	0.53	0.49
	Rep2	1.00	1.00	0.99	0.55	0.52	0.48
	Rep3	0.99	0.99	1.00	0.54	0.51	0.47
spt6l	Rep1	0.56	0.55	0.54	1.00	0.99	0.97
	Rep2	0.53	0.52	0.51	0.99	1.00	0.98
	Rep3	0.49	0.48	0.47	0.97	0.98	1.00

RNA-seq

		WT			*spt6l*		
		Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
WT	Rep1	1.00	0.96	0.95	0.52	0.57	0.51
	Rep2	0.96	1.00	0.98	0.53	0.58	0.51
	Rep3	0.95	0.98	1.00	0.55	0.58	0.52
spt6l	Rep1	0.52	0.53	0.55	1.00	0.98	0.99
	Rep2	0.57	0.58	0.58	0.97	1.00	0.97
	Rep3	0.51	0.51	0.52	0.99	0.97	1.00

24nt small RNA

Supplementary Figure 8