Organismal benefits of transcription speed control at gene boundaries

Leng, Xueyuan; Ivanov, Maxim; Kindgren, Peter; Malik, Indranil; Thieffry, Axel; Brodersen, Peter; Sandelin, Albin; Kaplan, Craig D.; Marquardt, Sebastian

Published in:
EMBO Reports

DOI:
10.15252/embr.201949315

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Organismal benefits of transcription speed control at gene boundaries

Xueyuan Leng¹, Maxim Ivanov¹, Peter Kindgren¹, Indranil Malik²†, Axel Thieffy³,⁴, Peter Brodersen⁴, Albin Sandelin³,⁴, Craig D Kaplan²,⁵ & Sebastian Marquardt¹,*,†

Abstract

RNA polymerase II (RNAPII) transcription is crucial for gene expression. RNAPII density peaks at gene boundaries, associating these key regions for gene expression control with limited RNAPII movement. The connections between RNAPII transcription speed and gene regulation in multicellular organisms are poorly understood. Here, we directly modulate RNAPII transcription speed by point mutations in the second largest subunit of RNAPII in Arabidopsis thaliana. A RNAPII mutation predicted to decelerate transcription is inviable, while accelerating RNAPII transcription confers phenotypes resembling auto-immunity. Nascent transcription profiling revealed that RNAPII complexes with accelerated transcription clear stalling sites at both gene ends, resulting in read-through transcription. The accelerated transcription mutant NRPB2-Y732F exhibits increased association with 5’ splice site (5’SS) intermediates and enhanced splicing efficiency. Our findings highlight potential advantages of RNAPII stalling through local reduction in transcription speed to optimize gene expression for the development of multicellular organisms.

Keywords NET-seq; speed; splicing; stalling; transcription

Introduction

A decisive step during gene expression is the conversion of the DNA sequences of a gene into pre-mRNA by RNA polymerase II (RNAPII) transcription. Profiles of RNAPII transcription across genes in eukaryotes revealed two main RNAPII localization peaks at gene boundaries, near gene transcription start sites (TSSs) and polyadenylation sites (PASs) [1]. At the 3’ end of genes, RNAPII peaks promote nascent RNA 3’ end processing and transcriptional termination in mammals [2,3]. The function of RNAPII peaks at promoter-proximal regions near TSSs is actively debated. On the one hand, “pause-release” of RNAPII can facilitate rapid induction of gene expression [4]; on the other hand, imaging of Drosophila and human RNAPII at promoter-proximal positions revealed rapid turnover, arguing against stable “pausing” of the same population of RNAPII complexes over time [5,6]. In metazoans, the negative elongation factor (NELF) complex promotes promoter-proximal pausing of RNAPII by limiting RNAPII mobility [7]. However, NELF is conspicuously absent in yeast and plants, which implies that many organisms use alternative mechanisms to stall RNAPII at promoter-proximal region (i.e., RNAPII stalling) [8]. In gene bodies, RNAPII accumulates at exon–intron boundaries and exhibits distinct accumulation profiles for exons with alternative splicing (AS) outcomes [9,10]. The efficiency of splicing may hence be coupled to the local speed of RNAPII elongation at exon–intron boundaries [11]. In summary, peaks of accumulated RNAPII represent sites with reduced RNAPII forward movement, which may facilitate the integration of cellular signals to control gene expression post-initiation by co-transcriptional RNA processing [12].

RNAPII forward movement depends on the dynamics of the trigger loop (TL), a central structure in the RNAPII active center [13–15]. In addition, RNAPII backtracking induced by weak RNA–DNA hybrids (i.e., nucleotide misincorporation) limits RNAPII forward movement [16–18]. A “gating tyrosine” in the RNAPII second largest subunit RPB2 (i.e., Y769 in budding yeast Rpb2) stacks with the first backtracked nucleotide and is proposed to prevent further backtracking [19] and is also positioned to interact with the TL when in its closed, catalysis-promoting state. Point mutations in budding yeast Rpb1 TL residues and Rpb2 TL-interacting residues alter the RNAPII elongation speed in vivo [20–24]. Such “kinetic RNAPII mutants” have informed greatly on the effects of altered transcription speed on gene expression and transcription-related phenotypes. For example, the budding yeast rpb2-P1018S slow transcription mutant (i.e., rpb2-10) promotes RNAPII arrest and reduces transcription processivity [25,26]. Moreover, kinetic RNAPII mutants displaying accelerated transcription favor the use of...
upstream TSSs, while mutants displaying slow transcription tend to use downstream TSSs [27]. Variations of transcription speed alter profiles of co-transcriptional chromatin signatures and of RNAPII C-terminal domain (CTD) phosphorylation that impact pre-mRNA processing [28–30]. These observations indicate a profound effect of RNAPII transcription elongation speed on gene expression. The important question of whether growth and differentiation programs in a multicellular organism can be executed when RNAPII carries kinetic point mutations remains largely unclear.

Here, we altered RNAPII transcription activity in Arabidopsis through point mutations in NRPB2, the second largest subunit of Arabidopsis RNAPII. A mutant accelerating RNAPII transcription triggered phenotypes consistent with auto-immunity, but was able to execute key steps of pattern formation and organogenesis. A mutation predicted to decrease RNAPII transcription speed was inviable. Nascent RNAPII transcription profiling revealed that the mutant accelerating transcription resulted in reduced RNAPII stalling at both gene boundaries. Our findings highlight mechanistic connections between the intrinsic speed of RNAPII and RNAPII stalling at both gene boundaries that coordinate gene expression in the context of a multicellular organism.

Results

Altering transcription activity of RNAPII by targeted mutagenesis of NRPB2

To alter the in vivo RNAPII transcription activity in whole plants, we generated point mutations in Arabidopsis RNAPII. The target residues were identified in Rpb2, the second largest budding yeast RNAPII subunit. The Rpb2 proline 1018 to serine substitution (rpb2-P1018S) represents the classic slow transcription mutant rpb2-10, and the tyrosine 769 to phenylalanine substitution (rpb2-Y769F) represents a mutation which might influence backtracking and TF function (Fig 1A) [19,31,32]. Sequence alignments identified P979S and Y732F in the highly conserved regions of RPB2, the second largest subunit of Arabidopsis RNAPII as the equivalent positions to budding yeast P1018S and Y769F, respectively (Fig 1B). We generated these point mutations in constructs carrying the genomic NRPB2 sequence fused to a C-terminal FLAG-tag driven by the endogenous NRPB2 promoter and integrated them into the nrbp2-2 null mutant background [33] (Fig EV1A). To investigate whether these point mutations affected NRPB2 protein accumulation, we performed Western blotting on FLAG-tagged NRPB2P979S-FLAG, NRPB2Y732F-FLAG, and wild-type NRPB2-FLAG (NRPB2WT-FLAG; Fig 1C). We identified several individual transformant lines with comparable steady-state protein levels; thus, any differences we detected in the characterization of these lines would have to be attributed to the effects of the point mutations on RNAPII activity.

The Arabidopsis nrbp2-2 null allele is female gametophytic lethal, but can be transmitted through the male germline with reduced transmission rate [33]. We could hence assay complementation of the gametophytic phenotypes to gain insights into the effects of RNAPII mutants. We assayed the transmission rate of the nrbp2-2 null allele in the plants carrying homozygous NRPB2WT, NRPB2P979S, or NRPB2Y732F transgenes in nrbp2-2+/− background (Fig EV1A). We would predict increased transmission rate of the nrbp2-2 allele if the gametophytic defects could be complemented. As predicted, NRPB2WT can fully (i.e., to the expected level of 50%) complement the transmission of nrbp2-2 compared to non-transformed controls (Fig 1D). Interestingly, NRPB2Y732F could almost fully complement nrbp2-2 transmission, while NRPB2P979S did not significantly increase transmission rate compared to non-transformed controls (Fig 1D). These data suggest that NRPB2P979S fails to provide the RNAPII activity necessary for germline development. Indeed, silique dissection revealed that the germline defects in NRPB2P979S nrbp2-2+/− were associated with reduced fertility and ovule abortion (Fig EV1B–D). Consistently, we identified plants homozygous for both NRPB2Y732F transgene and nrbp2-2 mutant (NRPB2Y732F+/+ nrbp2-2−/−) while NRPB2P979S+/+ nrbp2-2−/− genotype could not be recovered. Remarkably, when all RNAPII complexes carried the NRPB2Y732F mutation (i.e., NRPB2Y732F+/+ nrbp2-2−/−) we observed viable plant growth and development. These plants exhibited a dwarfed stature (Figs 1E and EV1E), but resembled Arabidopsis seedlings concerning basic patterning and organ formation. The dwarfed stature was reminiscent of mutants displaying auto-immunity, which is often associated with increased expression of pathogen-related (PR) genes [34]. Indeed, we detected elevated expression of PR1, PR2, and PR5 in NRPB2Y732F+/+ nrbp2-2−/− compared to NRPB2WT+/+ nrbp2-2−/− (Fig EV1F). These data highlight important roles of the ability to control the speed of RNAPII transcription during plant growth and development. In summary, Arabidopsis RNAPII harboring the NRPB2P979S point mutation failed to provide viable RNAPII activity during gametogenesis. However, the NRPB2Y732F mutation can partly rescue the germline defects in nrbp2-2 null mutants and allow plant growth and basic aspects of development.

NRPB2Y732F accelerates RNAPII transcription in vivo

To investigate the effect of NRPB2Y732F on RNAPII transcription speed, we first tested whether the equivalent rpb2-Y769F mutant in budding yeast classifies as a fast or slow RNAPII transcription mutant by assaying its sensitivity towards mycophenolic acid (MPA) and Mn2+ [35,36]. Budding yeast RNAPII mutants conferring enhanced catalytic activity (RNAPII fast mutants) are more sensitive towards Mn2+ than the RNAPII slow mutants [20]. Indeed, the rpb2-Y769F point mutation failed to provide fast catalytic rate and therefore putative elongation speed [27], where catalytically hyperactive RNAPII mutants exhibit an upstream shift in vivo. We detected mild sensitivity towards MPA compared to either single mutant (Fig EV2A). rpb2-Y769F mutant exhibited strong growth defects towards MPA and Mn2+ while we observed no effect for rpb2-P1018S (Fig EV2A). rpb2-Y769F thus shows a growth phenotype consistent with fast RNAPII transcription mutants [20]. Interestingly, the rpb2-Y769F/P1018S double mutant exhibited mild sensitivity towards MPA compared to either single mutant (Fig EV2A), consistent with a complementary effect on transcription speed as seen across many RNAPII active site mutations in budding yeast [27]. Primer extension analyses of alternative TSSs usage of the ADH1 gene represent an additional assay for RNAPII catalytic rate and therefore putative elongation speed [27], where catalytically hyperactive RNAPII mutants exhibit an upstream shift of TSS. In agreement with previously characterized fast RNAPII transcription mutants, rpb2-Y769F shifts the ADH1 TSS upstream compared to wild type or other Y769 substitutions (Fig EV2B). We
next tested the combinations of rpb2-Y769F with TL residue mutants previously demonstrated to alter RNAPII transcription speed. rpb2-Y769F was synthetically lethal with previously characterized fast RNAPII transcription mutants such as rpb1-E1103G, rpb1-E1103G, and rpb1-G1097D (Fig EV2C), suggesting that these combinations synergistically accelerated RNAPII transcription and supporting the interaction between Y769 and TL residues. Conversely, rpb2-Y769F suppressed the growth defect of previously characterized slow RNAPII transcription mutants such as rpb1-F1086S, rpb1-H1085Q, and rpb1-H1085Y [27,35] (Fig EV2C), suggesting compensatory effects on transcription speed when combining these “slow” mutations with rpb2-Y769F. In conclusion, our results characterized budding yeast rpb2-Y769F as a mutation conferring phenotypes consistent with hyperactive RNAPII mutants which increase RNAPII transcription speed.

To investigate the in vivo RNAPII transcription speed of Arabidopsis RNAPII carrying the NRPB2Y769F mutation, we developed an assay to monitor nascent RNAPII elongation after rapid transcription induction. To avoid time-consuming sample handling and processing issues associated with RNAPII chromatin

Figure 1. Altering transcription activity of RNAPII by targeted mutagenesis in NRPB2.

A Schematic drawing of Saccharomyces cerevisiae RNAPII transcription active center. Trigger loop is shown in blue. TL-interacting Rpb2 domain is shown in beige. Proline 1018 (P1018, green) and gating tyrosine 769 (Y769, red) are highlighted. The schematic drawing is based on PDB: 2e2h [15].

B Protein sequence alignment of RNAPII Rpb2 and P1018 regions in S. cerevisiae and Arabidopsis thaliana. P979S and Y732F are the yeast equivalent point mutations in Arabidopsis. The color scheme indicates conservation from variable (blue) to conserved (red).

C Detection of NRPB2WT-FLAG, NRPB2P979S-FLAG, and NRPB2Y732F-FLAG protein by Western blotting in NRPB2WT-FLAG Col-0, NRPB2P979S-FLAG Col-0, and NRPB2Y732F-FLAG Col-0 plants. Untagged NRPB2 (Col-0) was used as a negative control. Histone H3 was used as an internal control, and total protein level detected by stain-free blot was used as a loading control. Quantification was done by normalizing to the loading control and anti-H3 blot based on three independent replicates.

D Transmission rate of nrpb2-2 allele in nrpb2-2 WT line (n = 197) and nrpb2-2 WT lines combined with homozygous NRPB2p979S-FLAG+/+ (n = 280), NRPB2Y732F-FLAG+/+ (n = 240), and NRPB2p979S-FLAG+/+ (n = 240), respectively. Fisher’s exact test was used as a statistical test; three asterisks denote P < 0.001 between samples, and n.s. stands for not significant.

E Image of homozygous mutant nrpb2-2 fully complemented by NRPB2p979S-FLAG (top, NRPB2p979S-FLAG and rpb2-2 WT) and partially complemented by NRPB2Y732F-FLAG (bottom, NRPB2Y732F-FLAG and rpb2-2 WT). Plants were grown for 4 weeks in soil. Scale bars represent 1 cm.

Source data are available online for this figure.
immunoprecipitation from plants (RNAPII-ChIP) [25,37], we analyzed nascent RNA attached to RNAPII to monitor RNAPII elongation [38]. We identified three pathogen resistance-related Toll/interleukin receptor (TIR)-type NB-LRR genes AT4G19520, AT5G41740, and AT5G41750 [39,40] that are rapidly induced by flagellin 22 treatment. To monitor the “waves” of RNAPII elongation on these three genes after transcriptional induction, we performed a time course experiment during flagellin 22 treatment and determined the RNAPII signal by analyzing nascent RNA attached to RNAPII [38]. We chose NRPB2WT-FLAG+/+ Col-0 and NRPB2Y732F-FLAG+/+ Col-0 as material for this assay since we detected no differences in growth and immune response in this background. In brief, FLAG-tagged NRPB2WT and NRPB2Y732F proteins were immunoprecipitated by anti-FLAG antibody; RNAPII-associated RNA was purified and used in RT-qPCR analyses of three locations spanning these genes (Fig 2A). When gene induction is well synchronized, fast transcription is expected to show higher nascent RNA level in the gene body and towards the 3′ end of candidate genes during flagellin 22 treatment. We found that the candidate genes were rapidly induced by flagellin 22 treatment, as we detected an increase in nascent RNA level at probe 1 of these genes from 0 to 4 min after treatment (Figs 2B and C, and EV2E). Furthermore, data for the probe capturing RNAPII transcription shortly after induction (i.e., probe 1) suggest that these genes were induced with similar kinetics and to similar levels in NRPB2WT and NRPB2Y732F. Interestingly, we found that NRPB2Y732F showed higher nascent RNA level than NRPB2WT at probe 2 and probe 3 located further into the gene, from 3 min of flagellin 22 treatment onwards (Figs 2B and C, and EV2E). These data suggest that although wild-type RNAPII and mutant RNAPII were equally induced near the 5′ ends of genes, the NRPB2Y732F RNAPII reaches the 3′ ends of genes earlier than NRPB2WT, supporting faster RNAPII transcription in the NRPB2Y732F mutants. In summary, we detect evidence that the Arabidopsis NRPB2Y732F mutant exhibits accelerated RNAPII transcription in vivo.

**Accelerated RNAPII transcription promotes promoter-proximal RNAPII stalling**

To study the genome-wide effects of accelerated RNAPII transcription speed in NRPB2Y732F, we performed plant Native Elongating Transcript sequencing (plaNET-seq) to monitor nascent RNAPII transcription [8]. Two independent replicates of plaNET-seq were performed for NRPB2Y732F+/+ nrbp2-2−/− mutant and NRPB2WT+/+ nrbp2-2−/− control (Fig EV3A and B). Nascent RNA profiling in Arabidopsis revealed RNAPII stalling peaks near the beginning of transcription units in promoter-proximal regions. The positioning of the first nucleosome correlates well with the position of promoter-proximal RNAPII stalling in Arabidopsis [8]. To address the role of transcription speed in regulating promoter-proximal stalling, we investigated the RNAPII signal in promoter-proximal regions from plaNET-seq in NRPB2Y732F+/+ nrbp2-2−/− and NRPB2WT+/+ nrbp2-2−/−. Visual inspection suggested that NRPB2Y732F reduced peaks of RNAPII near the 5′ ends of genes when compared to NRPB2WT (Fig 3A). A metagene plot showing plaNET-seq RNAPII signal in a 1 kb region centered at the +1 nucleosomes [41] revealed that NRPB2Y732F reduced promoter-proximal RNAPII stalling centered at the +1 nucleosome position genome-wide compared to NRPB2WT (Fig 3B). The metagene-level reduction in RNAPII stalling in NRPB2Y732F was confirmed when the plaNET-seq signal was anchored at TSSs (Fig EV3C). To further quantify this effect, we calculated the RNAPII stalling index for well-expressed genes (plaNET-seq signal FPKM > 10, n = 6,596), which represents relative enrichment of RNAPII signal at promoter-proximal regions compared to the whole gene body. This analysis quantified a 35% reduction in the median value of RNAPII promoter-proximal stalling index in NRPB2Y732F compared to NRPB2WT (Fig 3C). These data illustrate that a restriction of RNAPII transcription speed contributes strongly to the formation of characteristic promoter-proximal RNAPII peaks.

**Accelerated transcription increases nascent RNAPII signal in gene bodies**

We observed increased RNAPII signals in NRPB2Y732F compared to NRPB2WT at intragenic positions downstream of promoter-proximal stalling sites (Fig EV3D). A metagene analysis of RNAPII activity across gene bodies confirmed this observation on a genome-wide scale (Fig 3D). Increased RNAPII signal in gene bodies could be reconciled by less RNAPII at promoter-proximal stalling regions in NRPB2Y732F compared to NRPB2WT. Consistently, increased nascent transcription in gene bodies in NRPB2Y732F correlated with increased plaNET-seq metagene profiles of exons and introns (Fig 3E and F). Interestingly, we detected an accumulation of RNAPII signal in gene bodies in the NRPB2Y732F mutant relative to NRPB2WT.
exonic plaNET-seq signal towards the 3’ end of exons in NRPB2<sub>y732F</sub> (Fig 3E). This effect was insensitive to the exon length (Fig EV3E–G). Exon–intron boundaries may thus trigger a pile-up of nascent RNAPII transcription when transcription is accelerated. In introns, accelerated RNAPII transcription amplifies nascent RNAPII signal compared to NRPB2<sub>WT</sub> and resulted in a uniform accumulation profile, which can be observed in metagene plots for introns of variable length genome-wide (Figs 3F, and EV3H–J). We next tested possible connections between increased intragenic nascent RNAPII signal and splicing regulation. However, the fast mutant showed increased signal over both constitutive and alternative exons and introns (Fig EV3K–N). In conclusion, accelerated RNAPII transcription in NRPB2<sub>y732F</sub> resulted in increased nascent RNAPII transcription in gene bodies.
Accelerated transcription enhances intron splicing and exon skipping

plA NET-seq co-purifies splicing intermediates due to co-transcriptional spliceosome association with RNAPII (Fig 4A). The splicing intermediates appear as single-nucleotide sharp peaks at 5’S’ splicing site (5’S’S) and 3’S’ splicing site (3’S’S) and thus can be distinguished from the nascent reads [10,42]. We detected an increased fraction of splicing intermediate reads corresponding to 5’S’S in plA NET-seq of NRPB2Y732F compared to NRPB2WT, while no obvious difference could be detected for 3’S’ splicing intermediates (Fig 4B). These data suggested an increased association of accelerated RNAPII transcription with splicing intermediates overlapping a 5’S’S. Since 5’S’ splicing intermediates are associated with the spliceosome, we predicted that higher RNAPII coverage in gene bodies could increase spliceosome association and perhaps enhance splicing in NRPB2Y732F. To test this idea, two independent replicates of RNA-seq were performed for NRPB2Y732F+/+ and NRPB2WT+/+ in seedlings (Fig EV4A). RNA-seq detects predominantly spliced transcripts with a characteristic signal intensity profile matching annotated exons. However, we noticed RNA-seq signal corresponding to some introns, presumably representing regulatory or poorly spliced introns (i.e., retained introns). Interestingly, initial visual

Figure 3. Accelerated RNAPII transcription reduces promoter-proximal RNAPII stalling and enhances RNAPII activity in gene body.
A plA NET-seq signal of RNAPII in the promoter-proximal region of AT1G70600 in NRPB2WT (red) and NRPB2Y732F (blue). The box plot shows the distribution of RNAPII signal (scaled to 0−1) in the promoter-proximal region (75 bp) for each sample. B Metagene profile of plA NET-seq mean signal of RNAPII in a 1 kb window centered at the +1 nucleosome in Arabidopsis genes (n = 25,474) in NRPB2WT (red) and NRPB2Y732F (blue). C RNAPII stalling index calculated for all the genes with plA NET-seq FPKM ≥ 10 in NRPB2WT (red) and NRPB2Y732F (blue). D, E Pla NET-seq signal for the mean signal of RNAPII across different regions, including: D whole genes (0.5−5 Kb scaled), whole genes 0.5−5 Kb scaled, scaled exons (100 bins), and Scaled introns (100 bins), respectively. F Metagene profile of plA NET-seq mean signal of RNAPII in introns (length from 50−300 bp) for each sample. G Metagene profile of plA NET-seq mean signal of RNAPII in exons (length from 50−300 bp, scaled to 100 bins, n = 73,925) in NRPB2WT (red) and NRPB2Y732F (blue). H Pla NET-seq signal for the mean signal of RNAPII across different regions, including: whole genes (0.5−5 Kb scaled), whole genes 0.5−5 Kb scaled, scaled exons (100 bins), and Scaled introns (100 bins), respectively.
inspection of several retained introns indicated that accelerated RNAPII transcription in NRPB2\textsubscript{Y732F} appeared to decrease intronic RNA-seq signal (Fig 4C). Strikingly, this finding is supported by a genome-wide decrease in the fraction of intronic RNA-seq signal across all genes (Fig 4D), suggesting a genome-wide trend of increased splicing efficiency in plants when RNAPII transcription is accelerated. A systematic genome-wide analysis identified 1,517 differentially expressed (DE) introns from the RNA-seq data of NRPB2\textsubscript{Y732F} compared to NRPB2\textsubscript{WT}. The majority (1,334 out of 1,517) of DE introns exhibit decreased fraction of intronic reads (Table EV2). We identified similar numbers of DE exons with increased or decreased expression in NRPB2\textsubscript{Y732F} compared to NRPB2\textsubscript{WT}, while we detected many more introns with decreased expression (Fig 4E). Quantification of DE exons revealed a small yet significant reduction in expression (Fig 4F) that we visualized for internal exons of the AT1G58060 and AT3G05680 genes (Fig 4G). In contrast, we detected a stronger decrease for DE introns in NRPB2\textsubscript{Y732F} mutant compared to NRPB2\textsubscript{WT} (Fig 4H).

We next tested alternative 5'SS and 3'SS usage (Fig EV4B) in the NRPB2\textsubscript{Y732F} mutant compared to NRPB2\textsubscript{WT} and found a trend to shift 5'SS upstream and 3'SS downstream (Fig EV4C–E). We note that a downstream shift of 3'SS is consistent with effects observed
in the splicing factor mutant ntr1 linked to increased transcription speed in Arabidopsis [16]. In summary, our RNA-seq data revealed multiple effects of altered RNAPII transcription on splicing in Arabidopsis. Our analyses highlighted reduced intron retention as the most notable effect of altered RNAPII activity in splicing in Arabidopsis revealed multiple effects of accelerated RNAPII transcription on splicing in Arabidopsis, with increased RNAPII transcription linked to increased transcription speed. To quantify this effect genome-wide, we determined the transcriptional read-through lengths in NRPB2WT and NRPB2Y732F. We used a statistical model which was based on empirical distributions of plaNET-seq tag counts in both genic and intergenic regions (see Materials and Methods). Strikingly, we observed that NRPB2Y732F extended transcriptional read-through genome-wide (FPKM > 5, n = 9,316; Fig 5C). We detected a 115 nt increase of median transcriptional read-through length in NRPB2WT compared to NRPB2WT (NRPB2Y732F, 649 nt; NRPB2WT, 534 nt; Fig 5D). NRPB2Y732F accelerating transcription speed thus reduces RNAPII termination efficiency and extends transcriptional read-through. The process of RNAPII termination is sensitive to the RNAPII active site and putative catalysis, consistent with a model where increased RNAPII speed alters kinetic competition between transcriptional stalling, termination, and further elongation.

Transcriptional read-through blurs the boundaries of transcription units, which could result in overlapping transcripts and potential gene expression conflicts. To investigate this, we focused on read-through transcription of tandemly oriented genes, where transcription read-through from upstream-located genes may invade downstream genes. RNAPII with accelerated transcription speed is expected to extend transcriptional read-through into the intergenic space (i.e., gaps) between the PAS of an upstream gene and the TSS of a downstream gene (PAS-TSS gaps). Indeed, NRPB2Y732F shows higher RNAPII signal than NRPB2WT in the second half of PAS-TSS gaps (n = 5,753) while RNAPII in NRPB2WT stalls downstream of PAS in the first half of PAS-TSS gaps (Figs 5E and EV5A). We further investigated plaNET-seq RNAPII signal in PAS-PAS gaps of paired genes facing each other in “tail-to-tail” orientation (n = 1,384). Also for this subset of genes, NRPB2Y732F lacked the characteristic RNAPII PAS-stalling in the first half of PAS-PAS gaps and showed significantly higher RNAPII signal in the second half of PAS-PAS gaps (Figs 5E and EV5B). These data suggest that accelerated transcription speed triggers transcriptional read-through genome-wide, resulting in overlapping transcripts and potential gene expression conflicts. In conclusion, our data highlight connections between reduced speed of RNAPII transcription at gene ends (i.e., PAS-stalling) and the termination of RNAPII transcription, linking the speed of transcription to spatial separation of plant transcription units.

### Figure 4. Analysis of alternative splicing in NRPB2WT and NRPB2Y732F mutant.

A A schematic illustration of the co-transcriptional RNAPII–spliceosome complex. plaNET-seq mainly detects splicing intermediates corresponding to 5' splicing site (5'SS) co-purified with engaged RNAPII complex.

B Bar charts showing the fractions of 5' and 3' splicing intermediate reads from plaNET-seq in NRPB2WT/+ nrbp2-2/- (NRPB2WT, blue) and NRPB2Y732F/+ nrbp2-2/-. Scale bars denote 0.5 kb.

C Genome browser snapshots illustrating enhanced intron splicing in NRPB2Y732F/+ nrbp2-2/- (NRPB2Y732F, red) compared to NRPB2WT/+ nrbp2-2/-. Scale bars denote 0.5 kb.

D The fraction of RNA-seq intronic reads calculated for all genes (n = 24,912) in NRPB2WT/+ nrbp2-2/- and NRPB2Y732F/+ nrbp2-2/-. Two-sided Mann–Whitney U-test. **** denotes P-value < 2.2e-16. The solid horizontal lines and box limits represent median, lower and upper quartiles of data values in each group. The upper and lower whiskers extend to the largest or smallest value, respectively, no further than 1.5*IQR from the relevant quartile.

E Differentially expressed (DE) exons and introns in NRPB2Y732F/+ nrbp2-2/- compared to NRPB2WT/+ nrbp2-2/- based on RNA-seq results. Numbers of DE exons and introns were shown in plot.

F Quantification (log2 fold change of FPKM from RNA-seq) of differentially expressed (DE) exons and non-DE exons in NRPB2Y732F/+ nrbp2-2/- compared to NRPB2WT/+ nrbp2-2/-. ** denotes P-value < 0.01 by Wilcoxon signed-rank test. The solid horizontal lines and box limits represent median, lower and upper quartiles of data values in each group. The upper and lower whiskers extend to the largest or smallest value, respectively, no further than 1.5*IQR from the relevant quartile.

G Genome browser snapshots illustrating enhanced exon skipping in NRPB2Y732F/+ nrbp2-2/- (NRPB2Y732F, red) compared to NRPB2WT/+ nrbp2-2/-(NRPB2WT, blue).

H Quantification (log2 fold change of FPKM from RNA-seq) of differentially expressed (DE) introns and non-DE introns in NRPB2Y732F/+ nrbp2-2/- compared to NRPB2WT/+ nrbp2-2/-. **** denotes P-value < 0.0001 by Wilcoxon signed-rank test. The solid horizontal lines and box limits represent median, lower and upper quartiles of data values in each group. The upper and lower whiskers extend to the largest or smallest value, respectively, no further than 1.5*IQR from the relevant quartile.

Source data are available online for this figure.
Our findings highlight molecular and organismal consequences of altered RNAPII elongation speed in a multicellular organism. The two main peaks of RNAPII localization in genomes at gene boundaries were depleted when transcription speed was accelerated (Fig 6A and B). Accelerated RNAPII transcription impacted gene expression after transcriptional initiation, through profound effects on splicing and transcriptional termination. Our data support that transcription speed control at gene boundaries is a key step in gene expression of multicellular organisms.

Discussion

RNAPII transcription speed and organismal development

While we succeeded in generation of viable plants carrying a fast RNAPII mutation, we were unable to obtain plants with a mutation in a conserved residue that reduced RNAPII transcription speed in yeast. This observation is reminiscent of embryonic lethality in mice through a point mutation in the largest RNAPII subunit that

Figure 5. Accelerated transcription reduces RNAPII stalling at 3′ gene ends and enhances transcriptional read-through downstream of PAS.

A plaNET-seq signal of RNAPII at 3′ end of AT2G21410 in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red). Arrows indicate the RNAPII signal peaks at PAS-stalling regions.

B Metagene profile of plaNET-seq mean signal of RNAPII in a 1 kb window centered at PAS (n = 24,448) in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red). The significance of differences of plaNET-seq signal in the region from PAS to +100 bins between NRPB2wt and NRPB2732F was assessed by two-sided Mann–Whitney U-test, P = 1.53e-06.

C Histogram of transcriptional read-through length (nt) from PAS of protein-coding gene (plaNET-seq FPKM > 5, n = 3,936) in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red).

D Box plot shows the RNAPII transcriptional read-through length from PAS of protein-coding genes (plaNET-seq, FPKM > 5, n = 3,936) called based on statistical model (see Materials and Methods) in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red). Medians of read-through length in NRPB2wt and NRPB2732F mutant are 534 and 649 nt. Two-sided Mann–Whitney U-test. *** denotes P = 9.9e-02. The solid horizontal lines and box limits represent median, lower and upper quartiles of data values in each group. The upper and lower whiskers extend to the largest or smallest value, respectively, no further than 1.5×IQR from the relevant quartile.

E Metagene plot of RNAPII signal by plaNET-seq anchored at both PAS of upstream genes and TSS of downstream genes for tandemly oriented genes (n = 5,753) in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red). Red arrow denotes the direction of transcriptional read-through. Pink dashed line rectangle indicates the region corresponding to the second half of PAS-TSS gaps along 5′–3′ direction.

F Metagene plot of RNAPII signal by plaNET-seq anchored at PASs of both upstream genes and downstream genes for gene pairs located in tail-to-tail orientation (n = 1,384) in NRPB2+/+ nrbp2-2/- (NRPB2wt, blue) and NRPB2732F/+ nrbp2-2/- (NRPB2732F, red). Red arrows denote the directions of transcriptional read-through from both PASs. Pink dashed line rectangles indicate the region corresponding to the second half of PAS-TSS gaps along 5′–3′ direction.
Promoter-proximal stalling represents a common feature of transcription throughout eukaryotic genomes [46]. The purpose of promoter-proximal RNAPII stalling is debated actively, yet a reduction in RNAPII transcription speed during stalling could be part of a checkpoint regulating gene expression. In organisms without NELF, for example, plants, promoter-proximal RNAPII stalling correlates with the position of the +1 nucleosome encountered by the transcription machinery [8]. Accelerated transcription in NRPB2<sup>y732F</sup> decreased promoter-proximal stalling and resulted in increased intragenic RNAPII transcription. Our data thus suggest that a reduction in RNAPII elongation speed near promoters facilitates the accumulation of promoter-proximal RNAPII peaks. In Arabidopsis, these peaks form independently of NELF and perhaps form through contributions by nucleosome barriers that correlate well with the peak position [8,47].

Possible advantages of promoter-proximal RNAPII stalling include a reduced response time to adjust gene expression to new environmental conditions. Instead of initiating the process of RNAPII transcription from recruitment and complex assembly, stalled RNAPII may represent pre-assembled and elongation competent RNAPII complexes waiting for signals to transcribe the full gene [1]. Defense signaling is crucial for plant fitness and regulated with fast temporal dynamics, perhaps achieved by a release of RNAPII into elongation from promoter-proximal stalling sites. If true, this would predict constitutive defense signaling when transcription is accelerated. Interestingly, NRPB2<sup>y732F</sup>/+ nrpb2-2/− plants resembled mutants with constitutively active defense signaling [34]. PR gene induction represents a diagnostic molecular hallmark of elevated defense signaling [48]. While alternative molecular explanations for stunted growth in NRPB2<sup>y732F</sup>/+ nrpb2-2/− may exist, for example, indirect effects, the induction of PR gene expression is consistent with an auto-immunity phenotype triggered by accelerated transcription. Our data thus provide a potential connection between plant defense signaling, promoter-proximal RNAPII stalling, and the speed of RNAPII transcription. In conclusion, these data imply that transcription speed limits at gene boundaries may benefit plants by avoiding constitutive defense signaling that triggers auto-immunity.

**Accelerated RNAPII transcription and RNA processing**

Our targeted introduction of candidate point mutations represents a direct approach to address mechanistic links between the speed of RNAPII transcription and RNA processing. Nevertheless, some molecular effects we reported could represent indirect effects caused by differences in growth and development between NRPB2<sup>y732F</sup>/+ nrpb2-2/− and NRPB2<sup>y732F</sup>/+ nrpb2-2/−. RNA-seq revealed that intron retention is reduced when RNAPII is accelerated; in other words, splicing efficiency of poorly spliced introns is increased. plaNET-seq data indicate that increased splicing efficiency is associated with the capture of splicing intermediates with terminal bases overlapping 5′SS, perhaps indicating that splicing of retained introns could be increased by promoting RNAPII binding to 5′SS. In conclusion, the speed of RNAPII transcription contributes to plant gene expression by modulating splicing efficiency, particularly at retained introns.

plaNET-seq data informed on transcriptional termination of RNAPII. Strikingly, we found a reduction in RNAPII peaks associated with termination when transcription is accelerated, and an increased distance of read-through transcription downstream of the PAS (Fig 5D). Read-through transcription triggered by elevated temperature has been reported in budding yeast and mammalian cell culture [37,49]. Extended read-through as observed in an accelerated RNAPII transcription mutant may have functional consequences on gene expression. The increased transcriptional read-through may result in gene expression defects for neighboring genes, for example, by transcriptional interference [50,51]. In summary, our data support the idea that a reduction in RNAPII transcription speed promotes RNAPII density peaks in genomes with functional consequences for the process of transcriptional termination.
Materials and Methods

Reagents and Tools Table

<table>
<thead>
<tr>
<th>Reagent or resource</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit Anti-Mouse Immunoglobulins/HRP</td>
<td>Dako</td>
<td>P0161; RRID:AB_2687969</td>
</tr>
<tr>
<td>Swine anti-rabbit Ig HRP antibody</td>
<td>Dako</td>
<td>P0217; RRID:AB_2728719</td>
</tr>
<tr>
<td>Anti-Histone H3 antibody</td>
<td>Abcam</td>
<td>Ab1791; RRID:AB_302613</td>
</tr>
<tr>
<td>Monoclonal ANTI-FLAG® M2 antibody produced in mouse</td>
<td>Sigma-Aldrich</td>
<td>F3165; RRID: AB_259529</td>
</tr>
<tr>
<td>Monoclonal ANTI-FLAG® M2 antibody produced in mouse</td>
<td>Sigma-Aldrich</td>
<td>F1804; RRID:AB_262044</td>
</tr>
<tr>
<td><strong>Bacterial and virus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> DH5α™ competent cells</td>
<td>Thermo Fisher</td>
<td>Cat no. 18265017</td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em> CV3101(PMP90)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Peptides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flagellin 22 peptides</td>
<td>Schafer-N</td>
<td>Peptide 40007</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>miRNeasy Mini Kit</td>
<td>QIAGEN</td>
<td>ID: 217004</td>
</tr>
<tr>
<td>SuperScript™ IV First-Strand Synthesis System</td>
<td>Invitrogen</td>
<td>Cat no.18091050</td>
</tr>
<tr>
<td>NEXTflex Small RNA-Seq Kit v3</td>
<td>Bioo Scientific</td>
<td>N/A</td>
</tr>
<tr>
<td>DNA High Sensitivity Kit</td>
<td>Agilent</td>
<td>5067-4626</td>
</tr>
<tr>
<td>4–15% Criterion™ TGX Stain-Free™ Protein Gel</td>
<td>Bio-Rad</td>
<td>Cat no. 5678084</td>
</tr>
<tr>
<td>TURBO DNA-free Kit</td>
<td>Thermo Fisher</td>
<td>Cat no. AM1907</td>
</tr>
<tr>
<td>Dynabeads™ Antibody Coupling Kit</td>
<td>Thermo Fisher</td>
<td>Cat no. 114311D</td>
</tr>
<tr>
<td>TruSeq RNA Library Prep Kit v2</td>
<td>Illuminia</td>
<td>RS-122-2001</td>
</tr>
<tr>
<td>GoTaq® qPCR Master Mix</td>
<td>Promega</td>
<td>Cat no. A6002</td>
</tr>
<tr>
<td>RNasey Plant Mini Kit</td>
<td>QIAGEN</td>
<td>ID: 74904</td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimental models: organisms/strains

| Arabidopsis thaliana Col-0 | N/A | N/A |
| NRPB2/nrpb2-2 mutant | Dr. Craig Pikaard | N/A |

Software and algorithms

| Software for bioinformatic analysis: see Materials and Methods | GitHub | https://github.com/Maxim-Ivanov/Leng_et_al_2019 |

Methods and Protocols

Plant material and growth conditions

The *Arabidopsis* mutant lines generated in this study were based on *A. thaliana* Columbia ecotype (Col-0) background. Generation of transgenic *Arabidopsis* plants was performed by *Agrobacterium*-mediated transformation as described [52]. NRPB2Y732F and NRPB2WT transgenes were first introduced to Col-0 and then crossed with nrpb2-2+/− mutant. NRPB2WT+/+ nrpb2-2/− and NRPB2Y732F+/+ nrpb2-2/− double mutants were screened by genotyping from F3 generation (Fig EV1A).

For in vitro growth, *Arabidopsis* seeds were surface-sterilized and placed on ½ MS media agar plates (1% sucrose) for 12 days, and seedlings were transferred into ½ MS liquid media (1% sucrose) and grew in flasks under 22°C with 16/8-h light/dark and 150 rpm shaking for 2 days. For growth on soil, *Arabidopsis* seeds were sowed on soil directly and undergo stratification in 4°C for 3 days before growth under 22°C with 16/8-h light/dark on soil.

Plasmid construction

The construction of vectors for *Agrobacterium*-mediated stable transformation was based on pEarleyGate 302 vector (pEG302). pEG302-NRPB2WT-FLAG construct was kindly provided by Craig Pikaard [33]. To generate pEG302-NRPB2Y732F-FLAG and pEG302-NRPB2WT-FLAG, pEG302-NRPB2WT-FLAG construct was linearized by DraII digestion and used as backbone in
isothermal assembly reactions; the inserts in isothermal assembly are partial genomic NRPB2 sequences containing Y732F (TAT to TTT) point mutation and P979S (CCG to TCG), respectively. Fragment containing DNA mutation for Y732F mutant was generated by overlapping PCR (primer pair 3089/3082) fusing two fragments generated by primer pair 3089/3467 and 3082/3466. By using similar strategy, fragment containing DNA mutation for P979S mutant was also generated by overlapping PCR (primer pair 3089/3082) fusing two fragments generated by primer pair 3089/3084 and 3082/3083. Isothermal assembly was performed subsequently to fusing two fragments generated by primer pair 3089/3084 and was also generated by overlapping PCR (primer pair 3089/3082) fusing two fragments generated by primer pair 3089/3467 and 3082/3466. By using similar strategy, fragment containing DNA mutation for P979S mutant was confirmed by DNA sequencing analysis. The primers used in plasmid construction are listed in Table EV1.

**Flagellin treatment**

Flagellin 22 (N-terminus acetylated) was synthesized by Schafer-N (https://schafer-n.com). For each replicate, flagellin 22 treatment was carried out by adding the 0.75 ml flagellin 22 solution (1 mg/ml in DMSO) to Arabidopsis seedlings from 50 µl seeds growing in 100 ml liquid MS media in a flask (3.3 µM as final concentration of flagellin 22). The treatment was set in 0-min (before treatment), 2-, 3-, and 4-min time course. Each experiment was performed in three independent replicates. After flagellin 22 treatment, the seedlings were flash-frozen in liquid nitrogen.

**Protein extraction and western blotting**

NRPB2WT-FLAG, NRPB2Y732F-FLAG, and NRPB2p979S-FLAG proteins were extracted from 2-week-old Arabidopsis seedlings from NRPB2WT/+ Col-0, NRPB2Y732F/+ Col-0, and NRPB2p979S/+ Col-0, respectively. Equal amounts of plant material were ground into a fine powder, and proteins were extracted in 2× extraction buffer (150 mM Tris–HCl pH 6.8; 5% SDS; 25% Glycerol; 0.025% Bromophenol blue; 0.1 mM DTT). Total proteins were separated by SDS–PAGE on precast 4–15% Criterion TGX stain-free protein gels (Bio-Rad) and transferred to PVDF membrane by Trans-Blot Turbo transfer system (Bio-Rad). 5% non-fat milk in PBS was used to block blotted membrane (30 min at room temperature). Anti-FLAG (Sigma F1804 or F3165) antibodies and anti-mouse HRP-conjugated secondary antibody (Dako P0217) were used as primary and secondary antibodies for the detection of FLAG-tagged NRPB2 proteins. Anti-H3 (abcam ab1791) antibody and anti-rabbit HRP-conjugated secondary antibody (Dako P0217) were used as primary and secondary antibodies for the detection of histone H3. The membrane was incubated with primary antibody overnight at 4°C with gentle rotation (final concentration 0.25 µg/ml in PBS). Membranes were washed with PBS and then incubated with secondary antibody (1:10,000 dilution in PBS) for 1 h at room temperature, followed by two-time washes with PBS (5 min each) and one-time wash with PBST (10 min). Chemiluminescent signals were detected using Super-Signal West Pico Chemiluminescent (Thermo Fisher Scientific) according to manufacturer’s instructions.

**Yeast strains, media, and primer extension analysis**

Yeast media are prepared as described [27]. For MPA and Mn²⁺ growth assay, MPA (final concentration 20 mg/ml) and MnCl₂ (15 mM) were supplemented to minimal SC-Leucine medium. The yeast RNAPII mutant strains were generated by site-directed mutagenesis as previously described [20]. TSS selection of ADH1 gene was assayed by primer extension analysis. In brief, corresponding yeast strains were grown in YPD media to mid-log phase; 30 µg of isolated yeast total RNA from each indicated strains was used in primer extension analysis exactly as previously described [20,27].

**Nascent RNA isolation**

Nascent RNA was isolated according to previously described protocol with minor changes [38]. NRPB2WT/+ Col-0, NRPB2Y732F/+ Col-0 seedlings from flagellin 22 treatment were ground into a fine powder. Nuclei were isolated and washed with HONDA buffer [0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 20 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 0.5% Triton X-100, 5 mM DTT, 1× EDTA-free Complete protease inhibitor (Roche)]. The nuclear fraction was digested by 600 U DNase I in 0.5 ml Lysis buffer [0.3 M NaCl, 20 mM Tris–HCl pH 7.5, 5 mM MgCl₂, 5 mM β-mercaptoethanol, 1× EDTA-free Complete protease inhibitor (Roche), 0.5% Tween-20, 10 µl RNase inhibitor (moloX GmbH, www.molox.de)] at 4°C for 20 min with shaking at 2,000 rpm on an Eppendorf ThermoMixer. The supernatant of a centrifugation (10,000 g for 10 min at 4°C) was recovered into a new tube and combined with Dynabeads M-270 (Invitrogen) coupled with anti-FLAG antibody (Sigma) for 2 h at 4°C with gentle rotation. Anti-FLAG antibody was coupled with Dynabeads according to the manufacturer’s instructions. After FLAG-IP, beads were washed 6 times using 0.5 ml Wash buffer [0.3 M NaCl, 20 mM Tris–HCl pH 7.5, 5 mM MgCl₂, 5 mM β-mercaptoethanol, 1× EDTA-free Complete protease inhibitor (Roche), and RNase inhibitor]. Bead-bound protein was eluted with 0.5 mg/ml 3xFLAG peptide (ApxBio) for 30 min twice at 4°C. RNA attached to immunoprecipitated proteins was isolated using QIAGEN miRNeasy Mini Kit according to manufacturer’s instructions. Western blot has been done as previously described [38] for input, unbound, and eluted fractions to monitor IP efficiency.

**Nascent RNA analysis**

Isolated nascent RNA was treated with Turbo DNase to remove DNA contamination following the manufacturer’s instruction (Ambion). Hundred nanograms of DNase-treated RNA was used for reverse transcription into cDNA by gene-specific primers following the manufacturer’s instruction of Superscript IV (Invitrogen) kit. Quantitative analysis of the generated cDNA was carried out by qPCR using the GoTaq qPCR Master Mix (Promega) and CFX384 Touch Real-Time PCR Detection System (Bio-Rad). Negative controls lacking the reverse transcriptase enzyme (–RT) were performed alongside all RT–qPCR experiments. qPCR expression level of each primer pair was calculated relative to the level of reference gene ACT2. All the primers used in RT and qPCR were summarized in Table EV1.

**plaNET-seq construction and sequencing**

Libraries for plaNET-seq were prepared as previously described [8]. Nascent RNA from NRPB2WT/+ nrpb2-2/– and NRPB2Y732F/+ nrpb2-2/– seedlings was used in plaNET-seq. In specific, the plaNET-seq libraries were constructed according to Bio Scientific’s NEXTflex Small RNA-Seq Kit v3 using a customized protocol. Two independent replicate libraries were constructed for each plant genotype. Approximately 100 ng RNA was used for each replicate.
The isolated RNA was ligated with 3'-linker and fragmented in alkaline solution (100 mM NaOCl pH 9.2, 2 mM EDTA). The fragmented RNA was cleaned up and subjected to T4 PNK treatment (20 U PNK, NEB) for 20 min at 37°C followed by re-annealing of RT primer (5'–GCCTTGGCAACCGAGAATTCCA-3'; 70°C, 5 min; 37°C, 30 min; and 25°C, 15 min). The RNA was then re-introduced to the manufacturer’s protocol at the adapter inactivation step. The final libraries were quantified with DNA High Sensitivity Kit on Agilent Bioanalyzer 2100 and then sequenced on the Illumina HiSeq 4000 platform in PE150 mode at Novogene (https://en.novogene.com).

**Total RNA extraction and RNA-seq**

Total RNA was extracted from 2-week-old NRPB2WT+/+ nrbp2-2/– and NRPB2Y732F+/+ nrbp2-2/– Arabidopsis seedlings using Plant RNeasy Mini Kit following manufacturer instructions (QiAGEN). Turbo DNase (Ambion) was used to treat extracted RNA using oligo-dT primers and Superscript IV (Invitrogen) as per manufacturer’s instructions. The poly(A)-enriched libraries for RNA-seq were constructed using Illumina TruSeq Sample Prep Kit v2 following the manufacturer’s protocol and quantified on Agilent 2100 Bioanalyzer. The sequencing was performed on Illumina HiSeq 4000 platform in PE100 mode.

**Bioinformatics**

All the supporting code for bioinformatics analysis is available at https://github.com/Maxim-Ivanov/Leng_et_al_2019.

Alignment and post-processing of plaNET-seq reads were done as previously described [8]. The first 4 bases of both R1 and R2 reads in plaNET-seq are unique molecular identifiers (UMIs). They were trimmed from read sequences and appended to read names using UMI-Tools v0.5.3. After UMI trimming, the 5’-terminal base of R2 corresponds to the 3’-end of original RNA molecule and thus denoted the genomic position of RNAPII active center. R2 reads were aligned to TAIR10 genome assembly using STAR v2.5.2b in transcriptome-guided mode with the following settings: –outSAMmultNmax 1 –alignEndsType Extend5p01Read1 –clip3pAdapterSeq GTACGGTCGACT. Ensembl Plants release 28 was used as the source of transcript annotation for alignment. The BAM files were sorted using Samtools v1.3.1. The following categories of reads were filtered out: (i) PCR duplicates (UMI-Tools); (ii) reads aligned within 100 bp from any rRNA, tRNA, snRNA, or snoRNA gene from Araport11 on either strand (BEDtools v2.27.0); and (iii) reads aligned with MAPQ < 10 (Samtools). The filtered BAM files were imported into R environment v3.5.1 using GenomicAlignments_1.18.1 library. The strand orientation of reads was flipped to restore strandedness of the original RNA molecules. 3’-terminal bases of flipped reads were found to overlap with donor or acceptor splice sites much more frequently than could be expected by chance. Such reads most likely represent splicing intermediates due to co-immunoprecipitation of the spliceosome together with FLAG-tagged RNAPII complexes. These reads were filtered out by overlap with the union of splice sites obtained from both Ensembl Plants 28 (TxDb.Athaliana.BioMart.plantsmart28 package) and Araport11 annotations. In addition, all split reads were removed as possible mature RNA contaminations. The remaining reads are expected to represent the nascent RNA population. Their genomic coverage was exported as strand-specific BigWig and bedgraph files using rtracklayer_1.42.2. For details on the alignment procedure, see 01-Alignment_of_plaNET-Seq_data.sh and 02-Postprocessing_of_-plaNET-Seq_data.R.

Araport11 annotation was used throughout all further steps of data analysis because it is more comprehensive in terms of non-coding transcripts than both TAIR10 and Ensembl Plants 28 annotations. We adjusted gene borders from Araport11 using coordinates of TSS and PAS tag clusters which were called using CAGEfightR package [53] from the available TSS-seq [50] and direct RNA-seq datasets [54,55], respectively. If multiple TSS or PAS tag clusters were connected to the same gene, the strongest of them was chosen as the new border. For details, see 03_Adjustment_of_Araport11_-gene_boundaries.

To draw metagene plots of plaNET-seq, we merged Bedgraph tracks of the two biological replicates of each genotype. The merged tracks were then normalized to 1 million reads in nuclear protein-coding genes. The X axes of metagene plots represent the genomic intervals of choice (whole genes, exons, introns, etc.) which were scaled to the defined number of bins. Intervals overlapping multiple annotated transcription units were excluded from consideration. Both introns and exons were trimmed by 5 bp each side prior to scaling to avoid possible artifacts. The Y axes show the sequencing coverage averaged between the genomic intervals. The shaded area in metagene plots represents normal-based standard error of mean of normalized plaNET-seq signal at each genomic bin. For details, see 08-Metagene_plots.R. The positions of nucleosomes in Arabidopsis were obtained from the PlantDHS database [56].

To calculate the read-through (RT) length, we considered transcribed genes (plaNET-seq FPKM in WT samples above 5). Genomic intervals for RT length estimation were defined from PAS of the analyzed gene to the nearest downstream TSS. Coordinates of TSS and PAS clusters were called from TSS-seq and direct RNA-seq datasets as described above. For each gene of interest, the empirical distribution of plaNET-seq tag counts in 100-bp sliding window was obtained (the “transcription” model). The “random” model corresponding to the un-transcribed state was represented by Poisson distribution where the rate parameter was estimated from plaNET-seq tag counts in intergenic regions. Then, plaNET-seq tags were counted in every 100-bp window moving in 10-bp steps along the candidate RT genomic intervals. For each window, the probability to observe at most this tag count under the gene-specific “transcription” model was divided by the probability to observe at least this tag count under the alternative “random” model. The start position of the first window where the probability ratio dropped below 1 was considered as the end of the read-through region. For details, see 04-Read-through_distance.R.

To calculate promoter-proximal RNAPII stalling index for each gene, we first found 100-bp windows with the highest plaNET-seq coverage within the interval (TSS – 100 bp, TSS + 300 bp). Center of this window was considered as the summit of promoter-proximal RNAPII peak. The stalling index was then calculated as the ratio of plaNET-seq coverage in this window vs the whole gene (normalized by gene width). To avoid statistical artifacts, genes shorter than 1 kb or having plaNET-seq FPKM below 1 were skipped from consideration. For details, see 05-Promoter-proximal_stalling_index.R.

RNA-seq reads were adapter- and quality-trimmed by TrimGalore v0.4.3 in paired-end mode and then aligned to TAIR10 by STAR v2.5.2b in local mode. Aligned reads with MAPQ below 10 were
removed by Samtools v1.3.1. BAM files were converted to unstranded Bedgraph and BigWig files using BEDtools genomcov v2.26.0 and kentUtils bedGraphToBigWig v4, respectively. The code was detailed in the section 06-Alignment_of_RNA-Seq_data.sh in the mentioned GitHub page.

Differentially expressed genes were called from RNA-seq data using DESeq2 [57]. Differentially expressed exons and introns were detected independently from the changes in gene expression level by DEXSeq [58]. Exons and introns were defined as disjoint exonic or intronic intervals, respectively, in Araport11. For details, see 07-Differential_expression.R.

To detect the differential usage of alternative 5’ and 3’ splice sites, transcript isoforms were first quantified by Cufflinks [59]. Then, the Cufflinks output was used to quantify the different AS events extracted from an Arabidopsis reference transcript dataset AtRTD2 [60] with SUPPA2 [61]. For details, see 10-SUPPA2_pipeline.sh and 11-Differential_AS.R scripts.

**Data availability**

The raw and processed plaNET-seq and RNA-seq data were deposited in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE133143.

**Expanded View for this article is available online.**

**Acknowledgements**

Research in the laboratory of S.M. is supported by a Hallas-Mäller Investigator award by the Novo Nordisk Foundation NNF15OC0014202 and a Copenhagen Plant Science Centre Young Investigator Starting grant. In addition, this project has received funding from the European Research Council (ERC) and the Marie Curie Actions under the European Union’s Horizon 2020 research and innovation program StC2017-757411 (S.M.) and MSCA-IF 730398 (P.K.). Research in the laboratory of C.D.K. is supported by NIH R01GM097260. We are grateful to Mary Gehring and Steve Buratowski for support during early stages of this project and Morgan Moeglein for excellent plant care. We are grateful to Craig Pikaard for sharing key reagents. We thank the members of the Marquardt lab for feedback on the manuscript.

**Author contributions**

Conceptualization: XL, CDK, SM; Methodology: MI, PK, SM; Investigation: XL, PK, IM, MI, SM; Formal analysis: MI, AT, AS; Data curation: MI; Writing—original draft: XL, SM; Writing—review and editing: XL, MI, CDK, PK, PB, AS, SM; Visualization, XL, MI; Resources: CDK, SM; Supervision: CDK, SM; Funding acquisition: SM.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**References**

phenotypic landscape of the RNA polymerase II trigger loop. PloS Genet 12: e1006321


© 2020 The Authors

License: This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.