BORDER proteins protect expression of neighboring genes by promoting 3’ Pol II pausing in plants

Xuhong Yu1,3, Pascal G.P. Martin1,2,3 & Scott D. Michaels1

Ensuring that one gene’s transcription does not inappropriately affect the expression of its neighbors is a fundamental challenge to gene regulation in a genomic context. In plants, which lack homologs of animal insulator proteins, the mechanisms that prevent transcriptional interference are not well understood. Here we show that BORDER proteins are enriched in intergenic regions and prevent interference between closely spaced genes on the same strand by promoting the 3’ pausing of RNA polymerase II at the upstream gene. In the absence of BORDER proteins, 3’ pausing associated with the upstream gene is reduced and shifts into the promoter region of the downstream gene. This is consistent with a model in which BORDER proteins inhibit transcriptional interference by preventing RNA polymerase from intruding into the promoters of downstream genes.
transcription of a gene does not occur in isolation but within the context of its genomic environment. The transcription of one gene has the potential to influence or interfere with that of its neighbors. Transcriptional interference (TI) can take many forms but is broadly defined as the direct negative impact of one gene’s transcription on a second gene that is located in cis. For example, if two genes are oriented in tandem on the same DNA strand, it is possible for the elongating RNA Polymerase II (Pol II) from the upstream gene to intrude into the promoter region of the downstream gene. This “promoter intrusion” has the potential to interfere with the binding of transcription factors, assembly of the preinitiation complex, and/or the positioning of nucleosomes at the promoter of the downstream gene. The potential for this type of TI may increase in genomes with higher gene density; shorter distances between genes would require more precise termination of upstream genes. Thus controlling elongation and termination at upstream genes may be key in preventing TI at downstream genes.

Accumulating evidence suggests important regulatory roles for Pol II pausing in shaping the transcriptome. An example of Pol II pausing seen in metazoans is the accumulation of transcriptionally engaged Pol II 30–50 bp downstream of the transcription start site (TSS). This promoter-proximal pausing is often seen at developmentally regulated genes, where it may facilitate their rapid activation, and is mediated by the DRB Sensitivity-Inducing Factor (DSIF) and the Negative Elongation Factor (NELF) complexes. Mapping of engaged Pol II in Arabidopsis thaliana and maize, in contrast, did not reveal patterns of Pol II accumulation in regions immediately downstream of TSSs. Thus, plants, which lack NELF homologs, do not appear to make significant use of promoter-proximal pausing. In a phenomenon known as 3′ Pol II pausing, however, a significant increase in Pol II is observed near the transcript end site (TES) of many genes. The molecular mechanisms that give rise to 3′ Pol II pausing in plants, as well as its biological significance, are unclear.

To better understand the role of 3′ pausing, we investigated a three-member family of putative negative transcription elongation factors from Arabidopsis, which we have named BDR (BDR1, BDR2, and BDR3) proteins. BDR proteins are enriched in intergenic regions and promote the 3′ pausing of Pol II for a large fraction of genes. This activity is especially important at closely spaced genes on the same strand (i.e., in tandem). In the bdr1,2,3 mutant, 3′ pausing is reduced at upstream genes and Pol II occupancy shifts into the promoter regions of the downstream genes. While expression of the upstream gene is unaffected in the bdr1,2,3 mutant, the shift in Pol II from the upstream gene into the promoter region of the downstream gene is coincident with reduced expression of the downstream gene. In this way, BDR proteins prevent TI between closely spaced tandem genes.

**Results**

**BDR proteins resemble transcriptional elongation factors.** BDR proteins form a three-member family in Arabidopsis (BDR1 = At5g25520, BDR2 = At5g11430, BDR3 = At2g25640). Each BDR protein contains a SPOC domain, which is found in the SPEN family of transcriptional repressors, and a transcription elongation factor IIS (TFIIS) central domain (Fig. 1a, Supplementary Fig. 1). TFIIS contains three domains (I, II/central, and III) and acts as a positive elongation factor. During elongation, RNA Pol II frequently backtracks, such that it is no longer positioned at the 3′ end of the growing transcript. To restart elongation, the central domain of TFIIS binds to RNA Pol II, while domain III stimulates cleavage of the nascent transcript, thus providing a new 3′ end for RNA Pol II. The fact that BDR proteins do not contain domain I or III suggests that the BDR proteins are unlikely to have TFIIS-like activity.

Proteins with similar domain organization are found outside plants, with fungal and animal proteins often including an additional N-terminal PHD domain. These include the mammalian proteins SPOCD1, PHF5, and DIDD1. The best characterized is the yeast protein BYPASS of Ess1 (Bye1), which contains a PHD domain in addition to its SPO and TFIIS central domains. Bye1 is thought to act as a negative elongation factor and binds to Pol II through its TFIIS central domain and to histone H3 trimethylated on lysine 4 (H3K4me3) through its PHD domain. Bye1 is enriched in the 5′ regions of genes and consistent with a role in repressing Pol II elongation. Pol II occupancy in the 5′ regions of genes is reduced in the bye1 mutant, whereas Pol II occupancy is increased in gene bodies.

To investigate the function of BDR proteins in Arabidopsis, we obtained T-DNA insertional mutants in BDR1 (bdr1-1), BDR2, (bdr2-1), and BDR3 (bdr3-1). Single mutants did not show clear phenotypes; however, the bdr1,2,3 triple mutant showed a short-root phenotype. Given the similarity between the BDR proteins and negative elongation factors, we speculated that the mutant phenotypes might be caused by increased transcriptional elongation. If this is the case, inhibiting elongation might attenuate the phenotype of the bdr1,2,3 mutant. To test this hypothesis, we grew seedlings in the presence of a chemical inhibitor of transcription elongation, 6-Azaauracil (6AU) and examined root growth. In contrast to wild type, which showed a reduction in root growth when grown on 6AU, root length was partially rescued in bdr1,2,3 mutant seedlings. The chemical inhibitor of transcription elongation, mycophenolic acid (MPA) had a slight negative effect on root growth in wild type (Fig. 1c) and Supplementary Data 1). We also tested a second chemical inhibitor of transcription elongation, mycophenolic acid (MPA) in contrast to wild type, which showed a reduction in root growth when grown on 6AU, root length was partially rescued in bdr1,2,3 mutant seedlings. The chemical inhibitor of transcription elongation, mycophenolic acid (MPA) had a slight negative effect on root growth in wild type (Fig. 1c) and Supplementary Data 1); however, bdr1,2,3 root length more than doubled when grown on MPA. These results suggest that the short-root phenotype may be due to increased transcriptional elongation in the bdr1,2,3 background.

**BDR proteins are enriched at gene borders.** We performed chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) to determine the localization of BDR1, BDR2, and BDR3 using MYC-tagged constructs driven by their respective endogenous promoters in the bdr1,2,3 background. All three constructs rescued the short-root phenotype of bdr1,2,3 (Supplementary Fig. 2A). ChIP-seq showed that BDR1 and BDR2 are mainly enriched at gene borders, with peak summits located a short distance upstream of TSSs and/or downstream of TESs (Fig. 2a, b). Because intergenic distances in Arabidopsis are relatively short (e.g., Fig. 2a), it is often not possible to unambiguously assign an intergenic peak to one of the two neighboring genes, but BDR peaks are nevertheless found between both converging and diverging gene pairs (Fig. 2a). In contrast to BDR1 and BDR2, which show roughly similar binding in TSS and TES regions, BDR3 showed a strong preference for TES binding (Fig. 2b). Among the three BDR proteins, BDR1 showed the highest ChIP-seq enrichment and BDR3 showed the lowest (Fig. 2b, note different y axis scales).

We defined 21,334, 11,997, and 12,178 peaks for BDR1, BDR2, and BDR3, respectively. Consistent with their greater amino acid sequence similarity (Supplementary Fig. 1), we found the greatest overlap in peaks between BDR1 and BDR2. Approximately 82% of BDR2 peaks overlapped with BDR1 peaks, whereas only 22% of BDR3 peaks overlapped with BDR1 (Fig. 2c). For all three BDR proteins, occupancy is correlated with the expression of the
nearest gene (Fig. 2b). BDR1 and BDR2 peaks were enriched in the intergenic regions, such as promoters and regions immediately downstream of the TES, as well as 5′ untranslated regions (5′UTRs; Fig. 2d). BDR3, in contrast, did not show enrichment in promoters or 5′UTRs but was enriched in exons, 3′UTRs, and regions immediately downstream of the TES. Because binding of BDR proteins is strongest near the TSS and/or TES, we examined occupancy in these regions in more detail. For each BDR protein, we identified sets of genes containing peaks within 300 bp of the TSS or TES and plotted the occupancy of the corresponding BDR protein (e.g., occupancy of BDR1 over TSS regions containing BDR1 peaks, occupancy of BDR2 over TSS regions containing BDR2 peaks, etc). In the TSS region, all three BDR proteins showed maximum occupancy slightly upstream of the TSS (Fig. 2e), with maxima of −87, −31, and −148 bp for BRD1, BDR2, and BDR3, respectively. BDR proteins showed less variation in binding position at the TES, with maxima of 106, 131, and 117 bp downstream of the TES for BRD1, BDR2, and BDR3, respectively. (Fig. 2f). BDR1 and BDR2 showed strong overlap in peaks at both the TSS and TES, whereas BDR3 showed stronger overlap with BDR1 than BDR2 at the TES than at the TSS (Fig. 2e, f).

BDR1 peaks are located in nucleosome-depleted, DNase-hypersensitive regions (Fig. 3a and Supplementary Fig. 3). Although less pronounced than for BDR1, regions with the highest occupancy for BDR2 and BDR3 also showed a preference for nucleosome-depleted regions (Supplementary Fig. 3).

Consistent with the correlation with gene expression (Fig. 2b), we found that the occupancy of all three BDR proteins also correlates with Pol II levels (Fig. 3a and Supplementary Fig. 3). Because BDR1, BDR2, and BDR3 show differences in binding in TSS and TES regions (Fig. 2b), we also examined the correlations between BDR proteins, Pol II, H3, and DNase-hypersensitive regions specifically at 250 bp regions immediately before the TSS, after the TSS, before the TES, and after the TES (Supplementary Fig. 4). These data show that some correlations are stronger in particular regions. For example, the correlation between BDR3, which shows relatively little binding near the TSS, with BDR1, BDR2, and Pol II is higher near the TES (Supplementary Fig. 4).

We also examined sequence conservation around BDR peaks. PhastCons25 examines sequence conservation between Arabidopsis and the genomes of 20 other angiosperms. In order to focus on the conservation of intergenic regions, sequences corresponding to annotated genes were removed. Because BDR peaks are preferentially found in nucleosome-depleted regions, we included other nucleosome-free regions, as well as random intergenic sequences, as controls. We found that sequence conservation was significantly higher at BDR peaks compared to surrounding intergenic sequences (Fig. 3b), with higher conservation observed for BRD1 and BDR2 peaks than for BDR3 peaks. We also searched for overrepresented motifs in BDR1 and BDR2 peaks, focusing on the 101 bp surrounding the peak center. Two motifs were identified that occurred more frequently in BDR1 and BDR2 peaks than in other intergenic regions (Fig. 3c, d). A TCP-like
motif was found in 44.9% of BDR1 peaks and an E-box motif was found in 7.02%. Both motifs were also enriched in BDR2 peaks (Fig. 3d). Interestingly, although BDR1 and BDR2 are enriched near both TSS and TES regions (Fig. 2b, e, f), these motifs only show enrichment near TSS sites (Fig. 3c). Because BDR proteins lack characterized DNA-binding motifs, it is likely that recruitment to chromatin depends on interactions with other factors. The result that enriched sequence motifs are found near the TSS, but not near the TES suggests that BDR proteins may be recruited to chromatin through multiple interactions/mechanisms, e.g.,
Fig. 2 Genome-wide localization of BDR proteins. **a** Browser track showing intergenic enrichment of BDR1, BDR2, BDR3, and Pol II. **b** Metagene profiles of BDR1::MYC, BDR2::MYC, and BDR3::MYC ChIP-seq signal in nine groups of genes defined by increasing mRNA expression levels in wild type. The average BDR ChIP-seq signal for each group (line) and the associated 95% confidence interval based on a Gaussian assumption (shade) are represented. Signal in gene bodies was averaged in bins of 1% of the gene size. FPKM fragments per kilobase per million aligned fragments. **c** Venn diagram showing the overlap between BDR1, BDR2, and BDR3 peaks. **d** Distribution of BDR ChIP-seq peaks in various classes of genomic features. Promoter regions and immediate downstream regions are defined as up to 300 bp upstream from the TSS or downstream of the TES, respectively. Intergenic regions are >300 bp from any gene. Asterisks indicate a significant enrichment compared to genome-wide distributions ($p < 0.002$). **e, f** Coverage of ChIP-seq signal for BDR1, BDR2, and BDR3 around the TSS (**e**) and TES (**f**). For each protein, genes were selected that contained peak summits <300 bp from their TSS (**e**) or TES (**f**). Venn diagrams illustrate the overlap between genes with BDR1, BDR2, and BDR3 peaks at their TSS (**e**) or TES (**f**).

Fig. 3 BDR peaks contain evolutionarily conserved TCP-like and E-box motifs. **a** Heatmap and metagene profiles (top) of ChIP-seq signals and DNAse-hypersensitive sites (DHS). Genes were sorted by total BDR1 signal around the TSS and TES; the top 15,000 genes are shown. **b** Sequence conservation across 20 angiosperms for intergenic regions around BDR1, BDR2, and BDR3 peak summits, nucleosome-free regions (NFR), or random regions. Average phastCons score (line) and 95% confidence intervals (shade) are shown. **c** Enrichment of TCP-like and E-box motifs in BDR1 and BDR2 summits, TSS, and TES regions. **d** Motifs identified in BDR1 and BDR2 summits.
interacting with DNA-binding proteins that recognize TCP-like and/or E-box motifs near the TSS, and other proteins, such as components of the transcription termination machinery, near the TES. The model that BDR proteins may be recruited to TSS and TES regions through separate mechanisms is also supported by asymmetric binding profile of BDR3, which shows much stronger affinity for TES regions than for TSS sites (Fig. 2b).

**BDR proteins promote 3′ Pol II pausing.** The potential role of BDR proteins as negative transcription elongation factors and their enrichment near the 3′ ends of genes suggests that they may play a role in 3′ pausing. We determined Pol II occupancy in wild type and bdr1,2,3 using antibodies recognizing Pol II, Serine 5 phosphorylated Pol II (S5P), and Pol II S2P. During transcription, Pol II undergoes a series of phosphorylation events, with Pol II S5P associated with initiation and Pol II S2P associated with elongation27. Consistent with this model, we observed that Pol II S2P signal increased through the body of the gene (Fig. 4a). S5P occupancy increased not only through the body of the gene but also showed a peak near the TSS and a depletion near the TES (Fig. 4a). Consistent with published ChIP-seq, GRO-seq, and pNET-seq studies10,11,28, all three antibodies showed 3′ Pol II accumulation just after the TES (Fig. 4a, red arrows), indicative of 3′ pausing.

We used ChIP-seq data from wild type and bdr1,2,3 to calculate a 3′ pausing index for Pol II (ratio of read densities from the region immediately downstream of the TES to those of the gene body, Fig. 4b). We first examined the relationship between 3′ pausing and gene expression. In wild type, the level of 3′ pausing was correlated with gene expression, with the most highly expressed genes having the highest levels of 3′ pausing (Fig. 4c). In bdr1,2,3, 3′ pausing was significantly reduced for nearly all combinations of antibody and gene expression group (Fig. 4c). Thus BDR proteins do indeed promote 3′ pausing for a large fraction of genes.

**BDR-protected genes occur in a specific genomic context.** We used RNA-seq analysis to identify three sets of genes whose expression is promoted or repressed by BDR proteins (i.e., show decreased or increased expression in bdr1,2,3 seedlings, respectively), as well as non-differentially expressed genes (Supplementary Fig. 5A). Interestingly, we found that BDR-promoted genes, which we will refer to as BDR-protected genes, preferentially occur in a specific genomic context (Fig. 5a and

---

Fig. 4 BDR proteins promote 3′ pausing and gene expression in a specific genomic context. **a** Pol II ChIP-seq coverage across expressed genes in Arabidopsis seedlings. 3′ pausing indicated by red arrows. **b** Calculation of a 3′ pausing index. **c** 3′ pausing indices for nine groups of genes defined by increasing mRNA expression levels in wild type. 3′ pausing is reduced in the bdr1,2,3 mutant, particularly at highly expressed genes. The centerline of boxplots is the median. The bounds of the box are the first and third quartiles (Q1 and Q3). Whiskers represent data range but are bounded to 1.5-fold the interquartile range (Q3–Q1); points outside this range are represented individually.
Supplementary Fig. 5B). In all, 74% of BDR-protected genes are on the same strand as their immediate upstream neighbor, compared to 50% for all genes in the genome (Fig. 5a and Supplementary Fig. 5B). No significant enrichment for orientation was found in the downstream neighbor of BDR-protected genes, although BDR-repressed genes showed a slight preference for having a downstream tandem neighbor (Supplementary Fig. 5B). In addition to orientation, we also examined the intergenic distances between BDR-protected genes and their upstream neighbors. The TES of the upstream gene is much closer to the TSS of BDR-protected genes (244 bp) compared to the genome-wide median of 859 bp (Fig. 5b). Finally, we examined the expression levels of tandem upstream genes. The tandem upstream neighbors of BDR-protected genes were more highly expressed (~2.5-fold higher) than the tandem upstream neighbors of non-differentially expressed controls or BDR-repressed genes (Fig. 5c). These tandem upstream genes were enriched for functions related to protein translation, subcellular targeting, and salt stress (Fig. 5c). In contrast to the BDR-protected genes themselves, the tandem upstream neighbors of BDR-protected genes were typically not differentially expressed in the bdr1,2,3 triple mutant. Thus BDR-protected genes preferentially occur a short distance downstream of a highly expressed gene on the same strand and BDR proteins are required to maintain the expression of the downstream gene but not the upstream neighbor.

**BDR proteins promote 3′ Pol II pausing.** Given that BDR-protected genes are generally located a short distance downstream of a highly expressed neighbor on the same strand, we speculated that 3′ pausing at the upstream gene might be important in protecting the downstream gene from TI. TI is broadly defined as the direct negative impact of one gene’s transcription on a second gene that is located in cis\(^1\). For example, it is possible for elongating Pol II from one gene to intrude into the promoter of a downstream gene on the same strand, disrupting its expression\(^2,3\).
To explore the model that BDR proteins might be important in promoting 3′ pausing at the upstream neighbors of BDR-protected genes, we examined the Pol II occupancy at the upstream neighbors of BDR-protected genes in wild type. We found that the upstream neighbors of BDR-protected genes have elevated 3′ pausing compared to a set of 1500 control genes with levels of expression similar to the upstream neighbors of BDR-protected genes (Fig. 6a and Supplementary Fig. 6). To determine whether the increased 3′ pausing is correlated with BDR proteins, we examined BDR protein occupancy at the upstream neighbors of BDR-protected genes. Compared to BDR-repressed genes or expression-matched control genes, BDR protein occupancy was
higher at the upstream neighbors of BDR-protected genes (Fig. 6b), particularly near the TES. Because BDR occupancy is correlated with gene expression levels and Pol II occupancy (Figs. 2b and 3a and Supplementary Fig. 3), we investigated whether Pol II occupancy could account for this enrichment. Even after normalization of BDR ChIP-seq coverage by Pol II occupancy, the enrichment in BDR1 and BDR2 occupancy in the intergenic region, reduction in the expression of the downstream gene in the bdr1,2,3 mutant (red arrows), and the reduction in 3′ paused Pol II at the upstream gene (blue arrows). The downstream shift in the position of 3′ pausing in bdr1,2,3 occurs preferentially at the upstream tandem neighbors of BDR-protected genes. The centerline of boxplots is the median. The bounds of the box are the first and third quartiles (Q1 and Q3). Whiskers represent data range but are bounded to 1.5-fold the interquartile range (Q3-Q1); points outside this range are omitted. f Upstream tandem neighbors of BDR-protected genes are enriched in gene loops36. Enrichment of gene loops in BDR-regulated genes and their neighbors. Statistics reflect the presence of loops in the circled gene in each context. p Values are shown for Fisher exact test with BH p value correction.

BDR proteins attenuate TI in response to photomorphogenesis. If reduced gene expression in bdr1,2,3 is the result of TI from upstream genes on the same strand, then increasing or decreasing the expression of upstream genes might exacerbate or relieve TI, respectively. To explore this possibility, we examined the changes in gene expression that occur at closely spaced tandem genes during photomorphogenesis, which results in the differential expression of a significant fraction of the genome. Wild-type and bdr1,2,3 seedlings were grown for 4 days in the dark. On the fifth day, seedlings were either maintained in darkness or transferred to white light for the final 2 or 4 h prior to RNA isolation (Fig. 7a). Gene expression changes in response to light were largely similar between bdr1,2,3 and wild type (Fig. 7b, c).

To look for evidence of increased TI when upstream genes are upregulated by light, we selected tandem genes with intergenic distances <600 bp, where expression of the upstream gene was similarly upregulated by light in both bdr1,2,3 and wild type. We then determined the ratio of expression levels (bdr1,2,3/wt) for the downstream genes under dark and light conditions. Consistent with our model, we found that expression of the downstream gene was significantly reduced in bdr1,2,3 upon upregulation of the upstream gene by light (Fig. 7d). We also found that TI could be relieved via the downregulation of upstream genes. Downstream genes that showed potential TI under dark conditions (i.e., reduced expression in the bdr1,2,3 mutant) showed a significant increase in expression when the upstream gene was downregulated by light (Fig. 7e). Taken together, these experiments show that the TI in bdr1,2,3 can be modulated by changing the expression of the upstream gene. Thus, in wild type, BDR proteins help to ensure the stable expression of downstream genes as their upstream tandem neighbors undergo light-regulated changes in gene expression.

We also observed that BDR proteins contribute to the rapid activation of light-regulated genes that have nearby upstream neighbors on the same strand, regardless of whether the upstream gene is light regulated. Among all genes that were light induced in wild type, we observed significantly reduced expression in bdr1,2,3 when the light-induced gene had an upstream neighbor on the same strand and <600 bp away (Fig. 7f). For example, biochemical pathway analysis of the genes showing reduced
induction in bdr1,2,3, showed a significant enrichment for genes encoding Calvin–Benson–Bassham (CBB) cycle enzymes (Supplementary Fig. 9), which uses ATP and NADPH created by photosynthesis to convert carbon dioxide and water into organic compounds\(^\text{37}\). Three CBB cycle genes are located a short distance from an upstream gene on the same strand (Fig. 7g). Even though the upstream genes are not induced by light, the downstream CBB cycle genes show attenuated induction by light in the absence of BDR proteins (Fig. 7g).

**Discussion**

TI between tandem genes was described in human alpha-globin genes >30 years ago\(^\text{38}\) and similar examples have been reported in yeast\(^\text{39,40}\), Drosophila\(^\text{41}\), or following a T-DNA insertion in Arabidopsis\(^\text{42}\). At the genome-wide scale, however, our understanding of how often and to what degree TI might shape the transcriptome is still limited. Examples from yeast suggest that transcription-dependent changes in nucleosome occupancy and histone marks at the promoter of the downstream gene may

---

\(^{37}\) Calvin, B.; Benson, G.A.; Bassham, D.C. (1966) Biochim. Biophys. Acta, 115, 140-156.

\(^{38}\) Tjian, R.; Maniatis, T. (1984) Cell, 37, 829-841.

\(^{39}\) Giraldo, J.; Pujol, J.; et al. (2008) Science, 320, 777-780.

\(^{40}\) Giraldo, J.; Pujol, J.; et al. (2008) Nature, 454, 489-493.

\(^{41}\) Giraldo, J.; Pujol, J.; et al. (2008) Science, 320, 777-780.

\(^{42}\) Giraldo, J.; Pujol, J.; et al. (2008) Nature, 454, 489-493.
contribute to TF 

This does not appear to be the case for BDR-protected genes, however, as our MNoase and histone ChIP-seq data indicate that the intergenic regions upstream from BDR-protected genes are nucleosome depleted. Rather, our data suggest that a reduction in 3' Pol II pausing and a shift in Pol II occupancy at the upstream gene are likely responsible for perturbing the function of the downstream promoter in bdr1,2,3 mutant. The precise mechanism by which BDR proteins promote 3' Pol II pausing is unclear; however, it is tempting to speculate that an increased Pol II elongation rate in bdr1,2,3 mutant might be responsible for a shift in Pol II termination site, as a downstream shift in termination has been observed using a "fast" elongating Pol II in human cells.

Although plants lack homologs of canonical animal insulator proteins, such as CTCF, the role of BDR proteins in ensuring that transcription of an upstream gene does not interfere with the expression of a closely spaced downstream neighbor can be thought of as a type of insulating activity. This suggests interesting parallels in the relationships between gene expression and chromatin organization in animals and Arabidopsis. In animals, it is common for enhancer elements to be located many kilobases away from the target gene. This creates a twofold problem of how to ensure that an enhancer element associates with/promotes the expression of the correct gene, while making sure that it does not affect the expression of other nearby genes. In animals, CTCF and cohesin help to solve both problems through the formation of loops/TADs, where sequences inside the loop are more likely to interact with each other than with sequences outside the loop. In this way, enhancer elements preferentially associate with genes inside the same loop and are "insulated" from genes outside the loop. Arabidopsis regulatory sequences, in contrast, are most often located near the promoter; examples of enhancer elements acting at a significant distance are rare. Thus there may be less need for CTCF-type insulators and large-scale TADs. The relatively compact genome of Arabidopsis, however, creates other problems, such as TI between closely spaced genes. Interestingly, part of the solution in plants may also involve chromatin loops. The enrichment of gene loops and BDR proteins in upstream tandem genes suggests that they may play a role in promoting 3' pausing and/or Pol II recycling thereby preventing TI with downstream neighbors. Taken together, these results indicate that BDR proteins inhibit TI by promoting 3' pausing at upstream genes, thereby protecting the promoter region of the downstream gene from invasion by upstream, terminating Pol II. It is interesting to note that, although 3' pausing is reduced at upstream genes in bdr1,2,3, the expression of the upstream genes themselves is usually not affected. This suggests that a gene's 3' pausing may be more important for protecting the expression of its neighbors than for its own expression. This type of an activity would be predicted to be particularly important in an organism, such as Arabidopsis, with relatively short intergenic regions.

Methods

Plant material and growth conditions. bdr1-1, SALK_142108C; bdr2-1, WISCED-SL0532H03; and bdr3-1, SALK_059905C were obtained from the Arabidopsis Biological Resource Center (Columbus, OH) and confirmed by Sanger sequencing. Plants were grown at 22 °C in long days (16-h light/8-h dark) under cool-white fluorescent light with a light intensity of approximately 125 μmol m−2 s−1.

Constructs. For epitope tagged constructs, the BDR1, BDR2, and BDR3 genomic DNAs without stop codons were transferred from pENTR to the destination vector pGWB16, which contains 4×MYC. Resulting constructs were used as templates to amplify BDR1:4×MYC, BDR2:4×MYC, and BDR3:4×MYC using primers that incorporate Sbf I and Spe I sites for (for BDR1: At5g25520-P1-sbf I-F cacctgcaggtcatatattatggggaaattcgagctc tctctttcccaaaaatttcaaaac+2701-pGWB16-myc-spe I-R actagtgatcggggaaattcgagctc tctctttcccaaaaatttcaaaac for BDR2: At3g11430-P1-sbf I-F cacctgcaggtcatatattatggggaaattcgagctc tctctttcccaaaaatttcaaaac +2701-pGWB16-myc-spe I-R for BDR3: At2g26540-P1-sbf I-F cacctgcaggtcatatattatggggaaattcgagctc tctctttcccaaaaatttcaaaac+2701-pGWB16-myc-spe I-R). The resulting fragments were first cloned into pENTR/D-TOPO and subcloned between Sbf I and Spe I sites in pmMXC309.

RNA expression analysis. For RNA-seq, total RNA was isolated from Arabidopsis seedlings using the Trizol reagent (Sigma) or Spectrum Plant Total RNA Kit (Sigma) or Plant/Fungi Total RNA Purification Kit (Norgen) following the manufacturer's instructions. The integrity of RNA samples was analyzed with Agilent Technologies 2200 Tape Station (Agilent Technologies). Input was quantified by using a Qubit RNA BR Assay Kit. RNA-seq Libraries were prepared from total RNA using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) and were sequenced on an Illumina NextSeq 500 at the Center of Genomics and Bioinformatics, Indiana University or Illumina HiSeq 2000 at the Genome Sequencing Facility in the Greehey Children's Cancer Research Institute of University Texas Health Science Center, San Antonio. All high-throughput sequencing data and corresponding experimental details are available in GEO SuperSeries GSE112443.

Chromatin immunoprecipitation followed by next-generation sequencing. Nuclei were isolated from cross-linked samples as described previously and resuspended in nuclei lysis buffer (50 mM Tris–HCl pH 8.0, 10 mM EDTA, 1% sodium dodecyl sulfate (SDS), 1 mM phenylmethylsulfonylfluoride (PMSF), 1% Plant Protease Inhibitors from Sigma). After fragmentation using a Covaris S200, the chromatin samples were diluted with ChIP dilution buffer (final concentration: 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl pH 8.0, 150 mM NaCl, 1 mM PMSF, 0.1% SDS, 1% Plant Protease Inhibitors, Sigma). The diluted chromatin samples were subjected to immunoprecipitation with antibodies (anti-MYC tag, clone 4A6, Millipore 05-724 (30 μg); Anti-RNA polymerase II CTD repeat YSPTSPS antibodies [BW916] Abcam ab117 (20 μg); Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody, Abcam ab5095 (30 μg); Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody Abcam ab5131 (30 μg); and control IgG Abcam ab18413 (20 μg)).

Native histone ChIP was performed as described previously using anti-Histone H3 Abcam ab1791 (10 μg). The ChIP libraries were prepared using the NEBNext Ultra® DNA Library Prep Kit (New England Biolabs) and then sequenced on a NextSeq 500 (Illumina)
at the Center of Genomics and Bioinformatics, Indiana University. All high-throughput sequencing data and corresponding experimental details are available in GEO SuperSeries GSE112443.

**RNA-seq computational analysis.** Two independent RNA-seq studies each with biological triplicates were performed in wild type and bdr1,2,3 triple mutant seedlings that were grown under standard conditions (GSE112440 and GSE112441). The second study (GSE112441) also included bdr1, bdr2, and bdr3 single mutant seedlings. Except when otherwise stated, study GSE112441 was used to compute figures presented in this manuscript. We systematically verified that consistent results were obtained with both the RNAseq studies. For GSE112440 (49 bp single-end reads sequenced on Illumina HiSeq 2000 instrument) and GSE112441 (2 x 43 bp paired-end reads sequenced on Illumina NextSeq 500 instrument), read alignments (for GSE112440: tophat2; for GSE112441: STAR) were performed using the exact mapping option with default parameters and a maximum insert size of 12 Kb. Duplicate fragments were removed with Picard 2.2.4 MarkDuplicates (http://broadinstitute.github.io/picard/). Samtools v 1.3 was used to keep only reads mapped in proper pairs with mapping quality (MapQ) > 20. For MNase-seq and ChiP-seq on histone modifications, we found that fragments <70 bp were enriched for background signal and fragments >250 bp for signal corresponding to dinucleosomes. Thus we only kept the corresponding reads to fragment sizes between 70 and 250 bp. Aligned reads were imported in R (v.3.3.2) to obtain coverages using Bioconductor v3.6.88. Coverage were normalized as fragment per 10 million fragments (FP10M) and exported to bigwig files with the rtracklayer package61. ChiP-seq peaks were detected using MACS2 2.1.062 in paired-end mode. Peaks located in blacklisted regions were removed. Annotation of peaks relative to genomic features were obtained using the ChiPpeakanno package63.

**ChiP-seq computational analysis.** All ChiP-seq samples were sequenced in paired-end mode on an Illumina NextSeq500 instrument (read length of 40, 43, or 155 bp, as specified in the corresponding GEO entries). Sequencing adapters were removed using Trimmomatic 0.33 in paired-end mode58 and reads were aligned to the *Arabidopsis* genome using Bowtie2 using the –draft parameter and a maximum insert size of 1 Kb. Duplicate fragments were removed with Picard 2.2.4 MarkDuplicates (http://broadinstitute.github.io/picard/). Samtools v1.3 was used to keep only reads mapped in proper pairs with mapping quality (MapQ) > 20. For MNase-seq and ChiP-seq on histone modifications, we found that fragments <70 bp were enriched for background signal and fragments >250 bp for signal corresponding to dinucleosomes. Thus we only kept the corresponding reads to fragment sizes between 70 and 250 bp. Aligned reads were imported in R (v.3.3.2) to obtain coverages using Bioconductor v3.6.88. Coverage were normalized as fragment per 10 million fragments (FP10M) and exported to bigwig files with the rtracklayer package61. ChiP-seq peaks were detected using MACS2 2.1.062 in paired-end mode. Peaks located in blacklisted regions were removed. Annotation of peaks relative to genomic features were obtained using the ChiPpeakanno package63.

**Average profiles and metagene plots.** Coverages (e.g., FP10M for ChiP-seq data, phastCons score, or annotation coverages) or normalized coverages (e.g., log2(BDR ChIP/WT control ChiP)) were directly used, without binning or smoothing, to produce average profiles centered on genomic features of interest (e.g., peak centers, TSS, or TES). After selecting a gene set of interest, the most extreme 0.01% of peaks were used to keep unique reads (sanitizing seq commands54), and gene-level read counting (featureCounts55 and differential change) (<0.25) for all comparisons (single gene set, upstream or downsteam, TSS, or TES). Multigene heatmaps were produced with the Enriched-Metagene heatmap plugin60, and gene size. For each group, we obtained metagene profiles using the ChiP-seq data for BDR1::MYC (left), BDR2::MYC (center), or BDR3::MYC (right) as described above.

**Multigene heatmaps.** Multigene heatmaps were produced with the Enriched-Metagene heatmap plugin60, for each gene set of interest, we obtained metagene profiles using the ChiP-seq data for BDR1::MYC (left), BDR2::MYC (center), or BDR3::MYC (right) as described above.
resulting position-weight matrices to scan all intergenic regions and evaluate the enrichment of the motifs under intergenic BDR1 peaks using a Fisher exact test.

Figure 6a. We calculated the strand-specific coverage of TAIR10 gene anotations and plotted the average coverage (line) and associated 95% CI (shade, normal assumption) around control. 3 Not all genes, and genes upregulated (BDR-repressed) or downregulated (BDR-protected) in the bdr1,2,3 mutant compared to wild-type plants. Genes significantly regulated in bdr1,2,3 mutant compared to wild type are illustrated in Supplementary Fig. 4A. The distance between the maximum values of the average coverage for wild type and bdr1,2,3 was evaluated using Paired Student’s t test with BH p value adjustment. Adjusted p values for all comparisons are shown.

Figure 5a. We selected the genes with an upstream gene on the same strand and a non-null read count in our RNA-seq study GSE112441. Then we plotted as boxplot the distribution of expression values (average log2(RPKM) from tripletic wild-type samples, GSE112441) for the upstream tandem genes and evaluated the significance of the differences between distributions using Mann–Whitney U test. Enrichment for GO biological processes among the upstream tandem neighbors of BDR-protected genes was evaluated with the goseq R package. Genes located upstream on the same strand of all expressed genes were used as the gene universe. Only categories with a p value < 0.001 are shown. Percentages of genes annotated with each GO category in the universe (gray) and in BDR-protected genes (green) are shown.

Figure 6a. For control genes (n = 1500, blue line) and for genes upregulated (n = 329, BDR-repressed, red line) or downregulated (n = 392, BDR-protected, green line) in the bdr1,2,3 triple mutant compared to wild-type plants, we selected the upstream genes located on the same strand and plotted their average coverage and associated 95% CI (shade) for BDR1:MYC, BDR2:MYC, and BDR3:MYC ChIP-seq fragments on their gene bodies and up to 2 Kbp on each side of gene borders.

Figure 6b. For genes downregulated in bdr1,2,3 (BDR-protected genes, same as for Fig. 6a), we identified the genes with an upstream gene on the same strand and plotted the average coverage (FP10M) of Pol II ChIP-seq data (antibodies as defined in Fig. 4a) obtained in wild type and the bdr1,2,3 triple mutant at the TES (±1 Kbp) of these upstream genes. The shade represents the corresponding 95% CI. The distance between the maximum values of the average coverage for wild type and bdr1,2,3 is indicated.

Figure 6d. We used the Gviz Bioconductor package to plot the coverage tracks for two genomic regions corresponding to genes downregulated in the bdr1,2,3 triple mutant (BDR-protected) and their upstream tandem neighbor. BDR1:MYC and BDR2:MYC ChIP-seq data (FP10M) are from GSE113059, BDR2:MYC (FP10M) is from GSE113072, Pol II ChIP-seq data (FP10M) in wild-type and the bdr1,2,3 triple mutant are from GSE113078, and RNA-seq data (RPKM) are averages from triplicates of wild type or bdr1,2,3 triple mutant samples from GSE112441.

Note that for RNA-seq data we used two different scales for the regulated genes (upregulated or downregulated) and for their upstream or downstream neighbors. The distance between the maximum values of the average coverage for wild type and bdr1,2,3 is indicated.

Figure 6f. Using the same data as for Fig. 6e, we assessed by Fisher exact test the enrichment of genes forming gene loops in genes regulated in the bdr1,2,3 mutant (upregulated or downregulated) and for their upstream or downstream neighbors located on the same or opposite strand. For each group of genes, the percentage of genes forming gene loops is indicated. p values from Fisher exact test were adjusted by the BH method.

Figure 7a. Experimental set-up of the photomorphogenesis experiment. The corresponding RNA-seq data are available in GEO series GSE112442.

Figure 7b. Principal component analysis was performed on the top 500 genes with the highest variance across samples using the DESeq2 plotPCA function.

Figure 7c. Intersections of the genes significantly regulated (DESeq2, FDR < 5%) in wild type and in the bdr1,2,3 triple mutant after 2 or 4 h light, compared to dark, were plotted as Venn diagrams using the VenneR package.

Figure 7d. RNA-seq data GSE112442 were analyzed with DESeq2 and identified differential expression induced by 2 or 4 h light compared to dark in wild type and the bdr1,2,3 triple mutant.

From all tandem gene pairs with an intergenic distance <600 bp, we selected those with an upstream gene that was upregulated (fold-change >2, BH-adjusted p value < 0.05, red, n = 278) or not differentially expressed (fold-change <1.5, BH-adjusted p value >0.2, blue, n = 251) at both time points (2 or 4 h) and in both wild type and the bdr1,2,3 triple mutant. We also removed genes showing evidence (p < 0.05) of a pre-existing TI (downregulation in bdr1,2,3 vs wt) under the dark condition because increased TI might be hard to detect for these genes (final number of genes, n = 214 controls and n = 231 genes with an upregulated upstream tandem neighbor). Then we plotted as boxplots the distribution of log2(bdr1,2,3/wt) values in each condition (dark, light 2 h, and light 4 h) for both groups of genes.

Figure 7e. From all tandem gene pairs with an intergenic distance <600 bp, we selected those with a downstream gene showing some evidence (p < 0.05) of TI (i.e., downregulated in bdr1,2,3 vs wt) under the dark condition and an upstream gene that was either downregulated (fold-change >1.5, BH-adjusted p value < 0.05, green, n = 148) in both wild-type and the bdr1,2,3 triple mutant at 2 or 4 h or were not differentially expressed (fold-change <1.5, BH-adjusted p value >0.1, blue, n = 761) in anageneity and at any time points. Then we plotted as boxplots the ratio of log2(bdr1,2,3/wt) values in each condition (dark, light 2 h, and light 4 h) for both groups of genes.

We used Wilcoxon signed-rank test to evaluate the change in log2(bdr1,2,3/wt) between the dark and light conditions. All p values < 0.05 are reported.

Figure 7f. Using RNA-seq data GSE112442, we selected all genes that were upregulated by light at 2 or 4 h in wild-type plants (DESeq2; fold-change >1.5, BH-adjusted p value < 0.05) and separated them by both orientation of their upstream gene (O: upstream gene is on the opposite strand, and S: upstream gene is on the same strand) and intergenic distance between their TSS and the upstream gene border (<600 bp, between 600 and 1200 bp or >1200 bp). For each group of genes under each condition (dark, light 2 h, and light 4 h), we plotted the distribution of the log2(bdr1,2,3/wt) for the downstream genes and compared genes with an upstream neighbor on the same strand (S) to having an upstream neighbor on the opposite strand (O) with Mann–Whitney U test with a BH p value adjustment. All adjusted p values < 0.05 are shown.

Figure 7g. The Gviz Bioconductor package was used to plot the coverage tracks for BDR1:MYC and BDR2:MYC (FP10M, GSE113059), Pol II ChIP-seq (FP10M, GSE113078), and RNA-seq data (average RPKM from triplicates in each group, GSE112442) for PCK1 (AT3G12780), GAPDH B subunit (AT1G42970), and FRAT (AT4G26520) and their respective upstream tandem gene neighbor.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Raw and processed ChIP-seq and RNA-seq data, along with detailed experimental and bioinformatic procedures, are provided in GEO Series GSE112443 and its subseries.

TAIR10 annotations were used for all analyses. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file.

Code availability
Author-generated computer codes and algorithms are available upon request.

Received: 24 January 2019 Accepted: 30 August 2019
Published online: 25 September 2019

References
1. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference—a crash course. Trends Genet. 21, 339–345 (2005).
2. Ard, R. & Allshire, R. C. Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res.* **44**, 16610–16630 (2016).

3. Hainer, S. J., Prunesi, J. A., Mitchell, R. D., Monteverde, R. M. & Martens, J. A. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* **25**, 29–40 (2011).

4. Venkatesh, S. et al. Set2 methylation of histone H3 lysine 36 suppresses alternative splicing in transcribed genes. *Nature* **489**, 452–455 (2012).

5. Venkatesh, S., Li, H., Gogol, M. M. & Workman, J. L. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat. Commun.* **7**, 13610 (2016).

6. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).

7. Gaertner, B. & Zeitlinger, J. RNA polymerase II pausing during development. *Development* **141**, 1179–1183 (2014).

8. Gilchrist, D. A. & Adelman, K. Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription. *Biochim. Biophys. Acta* **1819**, 700–706 (2012).

9. Levine, M. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 502–511 (2011).

10. Erhard, K. F. Jr., Talbot, J. E., Deans, N. C., McClish, A. E. & Hollick, J. B. Nascent transcription affected by RNA polymerase IV in Zea mays. *Genetics* **199**, 1107–1125 (2015).

11. Hessel, J., Duttke, S. H., Benner, C., Chory, J. & Nascent, R. N. A. sequencing reveals distinct features in plant transcription. *Proc. Natl Acad. Sci.* **113**, 12316–12321 (2016).

12. Chedin, S., Riva, M., Schultz, P., Sentenac, A. & Carles, C. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev.* **12**, 3857–3871 (1998).

13. Ariyoshi, M. & Schwob, J. W. A conserved structural motif reveals the essential transcriptional repression function of Spn proteins and their role in developmental signaling. *Genes Dev.* **17**, 1909–1920 (2003).

14. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II-TFIIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).

15. Sims, R. J. 3rd, Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18**, 2437–2468 (2004).

16. Selth, L. A., Sigurdsson, S. & Svejstrup, J. Q. Transcript elongation by RNA polymerase II. *Annu. Rev. Biochem.* **79**, 271–293 (2010).

17. Kinkelin, K. et al. Structures of RNA polymerase II complexes with Byc1, a chromatin-binding PHF3/DIDO homologe. *Proc. Natl Acad. Sci. USA* **110**, 15272–15282 (2013).

18. Fischer, U. et al. PHF3 expression is frequently reduced in glioma. *Cytogenet. Cell Genet.* **94**, 131–136 (2001).

19. Tencer, A. H. et al. A unique pH-dependent recognition of methylated histone H3K4 by PPS and DIDO. *Structure* **25**, 1530–1539 e3 (2017).

20. Li, J., Zhao, Z., Ho, J., Liu, Y. & Li, Z. SPOC1 promotes cell proliferation and inhibits cell apoptosis in human osteosarcoma. *Med. Mol. Path. Rep.* **17**, 3218–3225 (2018).

21. Liu, Q. et al. SPOC1 promotes the proliferation and metastasis of glioma cells by up-regulating PTX3. *Ann. J. Cancer Res.* **8**, 624–635 (2018).

22. Harlen, K. M. & Churchman, L. S. Subgenic Pol II interactomes identify region-specific transcription elongation regulators. *Mol. Syst. Biol.* **13**, 900 (2017).

23. Pinskaya, M. et al. PHF3-like domains of the byc1 transcription factor determine its multivalent genomic distribution. *PLoS ONE* **9**, e102464 (2014).

24. Exinger, F. & Lacroute, F. 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. *Curr. Genet.* **5**, e102464 (2014).

25. Fong, N. et al. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* **60**, 256–267 (2015).

26. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

27. Law, C. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 50–58 (2013).

28. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

29. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

30. Love, M. L., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

31. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a versatile tool for trimming Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **10**, 377–378 (2013).

33. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for performing genomically distributed data analysis. *Bioinformatics* **25**, 1215–1220 (2009).

34. Andronache, A. & Wang, L. G. & He, Q. Y. ChiPseeker: an R/Bioconductor package for ChiP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
Gu, Z., Eils, R., Schlesner, M. & Ishaque, N. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics* **19**, 234 (2018).

Brown, G. GreyListChIP: grey lists–mask artefact regions based on ChIP inputs. *R* package version 1.10.0. (Bioconductor, 2016).

Hahne, F. L. in *Statistical Genomics: Methods and Protocols* (eds E, Mathé, E. & Davis, S.) 335–351 (Springer New York, New York, 2016).

Swinton, J. Vennerable: Venn and Euler area-proportional diagrams. *R* package version 3.1.

Zhang, W., Zhang, T., Wu, Y. & Jiang, J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell* **24**, 2719–2731 (2012).

Thomas-Chollier, M. et al. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat. Protoc.* **7**, 1551–1568 (2012).

### Acknowledgements

Next-generation sequencing was performed at the Indiana University Center for Genomics and Bioinformatics. We thank Mathilde Gorieu for technical help during the photomorphogenesis experiment. We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources. This work was supported by a grant to S.D.M. from the National Institutes of Health (GM075060). P.G.P.M. received the support of the EU in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills+ fellowship (under grant agreement no. 609398).

### Author contributions

All authors participated in the design of experiments. X.Y. created genetic materials, X.Y. and P.G.P.M. created sequencing libraries, and P.G.P.M. performed bioinformatic analyses.

### Additional information

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-019-12328-w.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Peer review information** Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019