Chr1:28979275-TDNA

>SRR14850978.23897795
CAATTGCAGGAAAAGTAGCGGATTCTCGCGTGGTGCATCCATGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCG
GCCGAGTTG CTTACATAATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCA
>AT1G51190
CAATTGCAGGAAAAGTAGCGGATTCTCGCGTGGTGCATCCATGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCG
GCCGAGTTG **CT**GGAAACAAAGATCTCTACTTGGGAACATTCA
>pAC161
AATTCAGTGTAGCCCATT**AC** CTTACATAATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTG

Chr1:28979291-TDNA

>SRR14850977.2319171
GCGGATTCTCGCGTGGTGCATCCATGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCGGCCGAGTTGCTGGAAA
CAAAGATCT CTTACATAATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAA
>AT1G51190
GCGGATTCTCGCGTGGTGCATCCATGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCGGCCGAGTTGCTGGAAA
CAAAGATCT **CT**ACTTGGGAACATTCA
>pAC161
AATTCAGTGTAGCCCATT**AC** CTTACATAATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTG

Chr1:28979571-TDNA

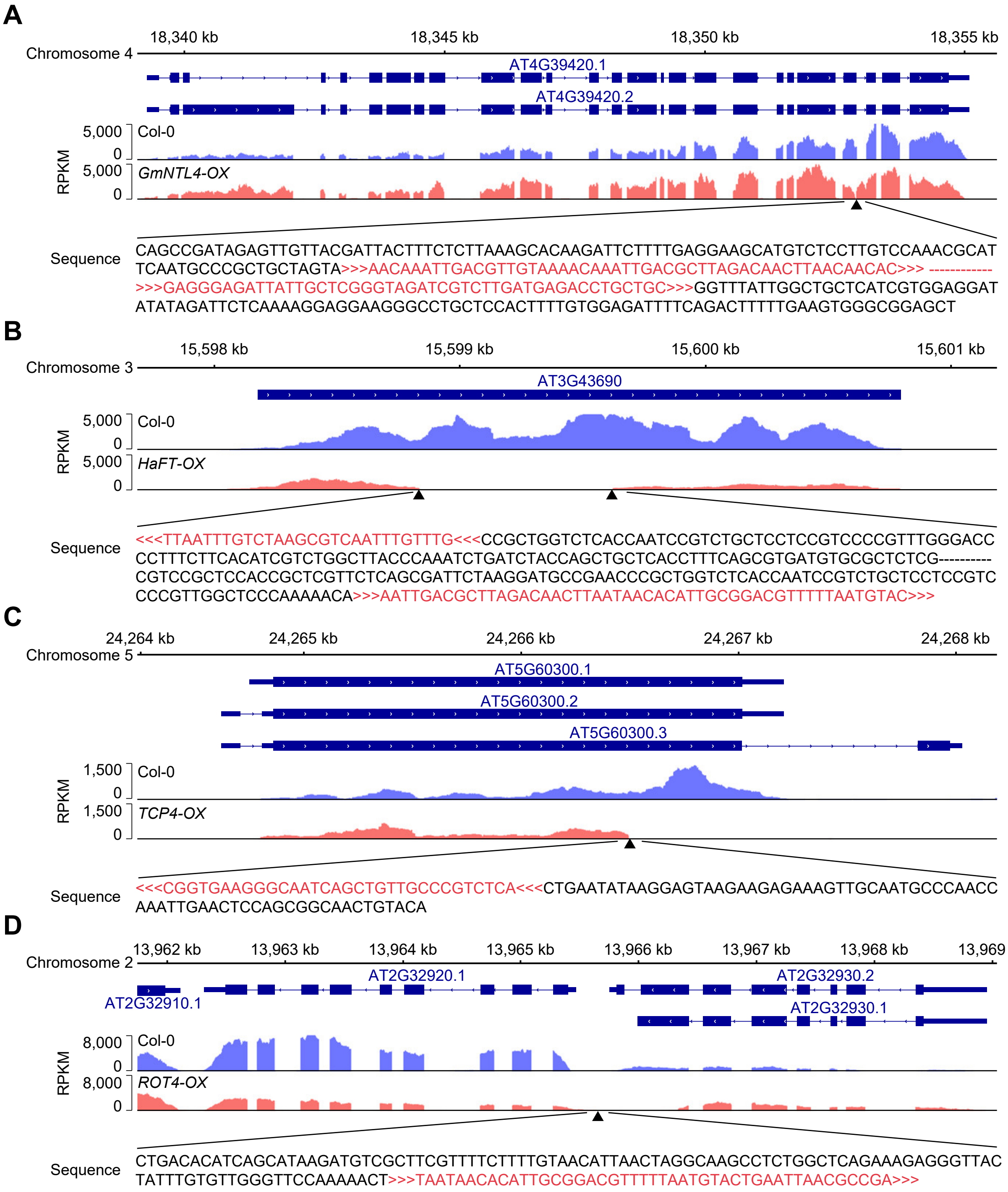
>SRR14850977.60730996
TGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCGGCCGAGTTGCTGGAAACAAAGATCTCTACTTGGGAACATTC
A GCAGTCTGTCAACTCGGCGTCAATTTGTCGGCCACTATACGATAGTTGCGCAAAT
>AT1G51190
TGTATCGTGGAGTAACAAGGCATCATCAACATGGAAGATGGCAAGCAAGGATCGGCCGAGTTGCTGGAAACAAAGATCTCTACTTGGGAACATTC
A **GT**AAGAATAAATATTAA
>pAC161
CACTCAAATGCT**AG** GCAGTCTGTCAACTCGGCGTCAATTTGTCGGCCACTATACGATAGTTGCGCAAAT

Chr3:7302676-TDNA

>SRR14850978.50595879
GGGAACTTTTAGCACTGAGGAAGAAGCAGCAGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGAGATCAA
CCG TGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGC
>AT3G20840
GCACTGAGGAAGAAGCAGCAGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGAGATCAACCG
GTACGACGTGAAAGCCATT
>pROK2
GTTTTTCTTTTCACC**AG** TGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGC

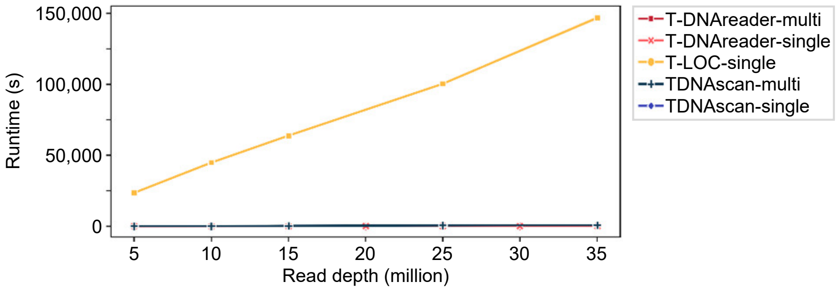
Supplementary Figure S1. Examples of T-DNAreader-identified T-DNA insertion sites (TISs) from the *plt1-21 plt2-21* mutant.

The RNA-seq chimeric reads containing both plant genomic and T-DNA sequences are shown. Original plant genome and T-DNA sequences near the TISs are also shown for comparison. The genomic and T-DNA sequences that overlap with the chimeric reads are colored blue and red, respectively. Splicing junction sequences of GT or AG are highlighted in green.

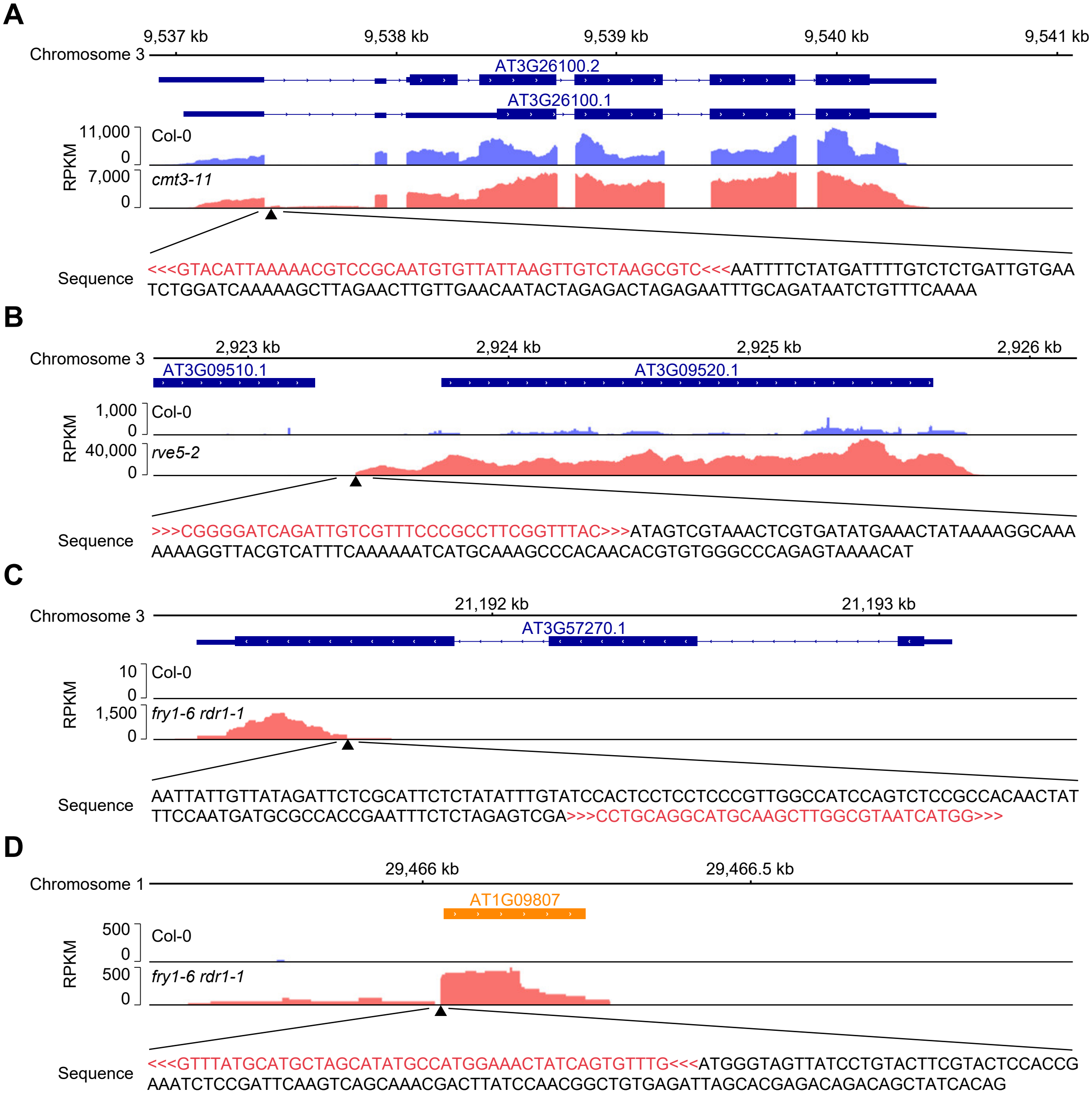


Supplementary Figure S2. Examples of T-DNAreader-identified T-DNA insertion sites (TISs) from RNA-seq data of transgenic overexpression lines.

A–D RNA-seq coverage plots at T-DNAreader-identified TISs in the wild type (Col-0) and transgenic overexpression lines. The RNA-seq coverage was calculated as the number of reads per kilobase of transcript per million mapped reads (RPKM). The black triangles indicate the T-DNAreader-identified TISs. The T-DNA sequences are colored red, and the orientation of the T-DNA insertion is indicated.

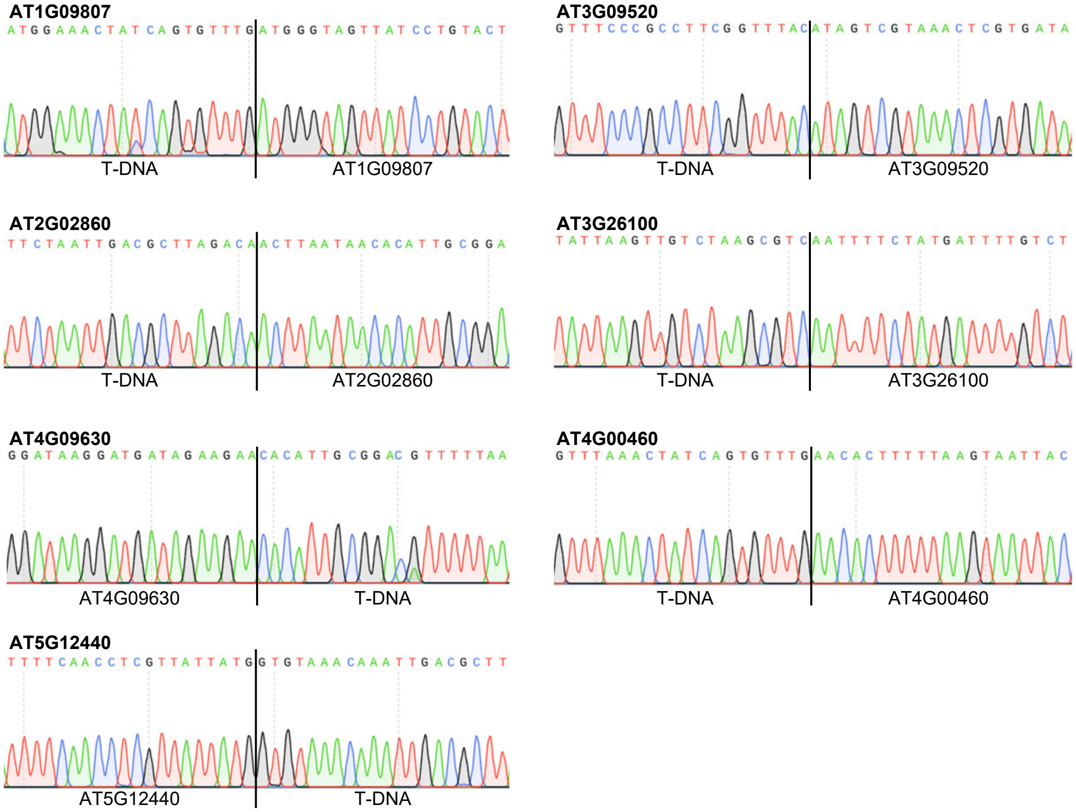


Supplementary Figure S3. Runtime comparison between T-DNAreader, T-LOC, and TDNAscan. A runtime comparison between T-DNAreader, T-LOC, and TDNAscan was conducted using subsampled data from SRR19019339. T-DNAreader was implemented with a single thread (T-DNAreader-single) and with two threads (T-DNAreader-multi). For T-LOC, the runtime was estimated using T-LOC-single with single control data. TDNAscan was implemented with a single thread (TDNAscan-single) and with eight threads (TDNAscan-multi).



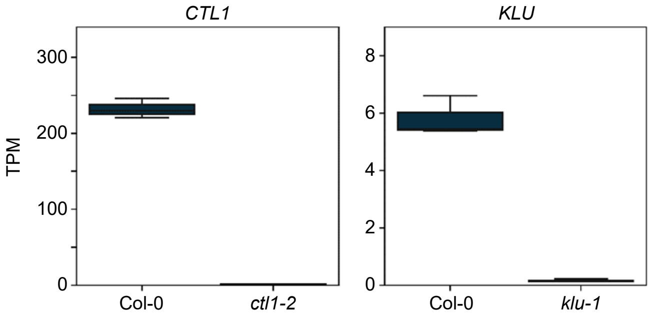
Supplementary Figure S4. Misregulation of RNA splicing or gene transcription at the previously uncharacterized T-DNA insertion sites (TISs).

A–D RNA-seq coverage plots at the previously unannotated TISs in the wild type (Col-0) and T-DNA insertion mutants. The RNA-seq coverage was calculated as the number of reads per kilobase of transcript per million mapped reads (RPKM). The black triangles indicate the T-DNAreader-identified TISs. The T-DNA sequences are colored red, and the orientation of the T-DNA insertion is indicated.



Supplementary Figure S5. Sequencing results of amplified chimeric fragments at the previously unannotated T-DNA insertion sites (TISs).

The Sanger sequences of the amplified chimeric fragments at the previously unannotated TISs in the *Arabidopsis* genes *AT1G09807*, *AT2G02860*, *AT3G09520*, *AT3G26100*, *AT4G00460*, *AT4G09630*, and *AT5G12440* are displayed.



Supplementary Figure S6. Expression levels of the *CTL1* and *KLU* genes in the T-DNA insertion mutants.

Expression levels of the *CTL1* and *KLU* genes in the wild type (Col-0) and the T-DNA insertion mutants *ctl1-2* and *klu-1*. The *ctl1-2* and *klu-1* mutants contain T-DNA insertions within their promoter and 5′-untranslated regions, respectively. Log$_2$(transcripts per million (TPM) + 1) was used for the quantification of the expression levels.