



## LARGE-SCALE BIOLOGY ARTICLE

# Genome-Wide Association Analyses Reveal the Importance of Alternative Splicing in Diversifying Gene Function and Regulating Phenotypic Variation in Maize<sup>[OPEN]</sup>

Qiuyue Chen,<sup>a,1</sup> Yingjia Han,<sup>a,b,1</sup> Haijun Liu,<sup>c,1</sup> Xufeng Wang,<sup>a</sup> Jiamin Sun,<sup>c</sup> Binghao Zhao,<sup>a,b</sup> Weiya Li,<sup>a,b</sup> Jinge Tian,<sup>a</sup> Yameng Liang,<sup>a</sup> Jianbing Yan,<sup>c</sup> Xiaohong Yang,<sup>a,b,2</sup> and Feng Tian<sup>a,2</sup>

<sup>a</sup>National Maize Improvement Center of China, MOA Key Laboratory of Maize Biology, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

<sup>b</sup>State Key Laboratory of Plant Physiology and Biochemistry, China Agricultural University, Beijing 100193, China

<sup>c</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

ORCID IDs: 0000-0002-3304-8321 (Q.C.); 0000-0002-3397-9361 (Y.H.); 0000-0001-7717-893X (H.L.); 0000-0001-9345-827X (X.W.); 0000-0001-9903-0629 (J.S.); 0000-0002-9604-300X (B.Z.); 0000-0002-6106-8530 (W.L.); 0000-0002-0868-3163 (J.T.); 0000-0002-0104-0413 (Y.L.); 0000-0001-8650-7811 (J.Y.); 0000-0001-6438-948X (X.Y.); 0000-0003-3552-4536 (F.T.)

**Alternative splicing (AS) enhances transcriptome diversity and plays important roles in regulating plant processes. Although widespread natural variation in AS has been observed in plants, how AS is regulated and contribute to phenotypic variation is poorly understood. Here, we report a population-level transcriptome assembly and genome-wide association study to identify splicing quantitative trait loci (sQTLs) in developing maize (*Zea mays*) kernels from 368 inbred lines. We detected 19,554 unique sQTLs for 6570 genes. Most sQTLs showed small isoform usage changes without involving major isoform switching between genotypes. The sQTL-affected isoforms tend to display distinct protein functions. We demonstrate that non-sense-mediated mRNA decay, microRNA-mediated regulation, and small interfering peptide-mediated peptide interference are frequently involved in sQTL regulation. The natural variation in AS and overall mRNA level appears to be independently regulated with different *cis*-sequences preferentially used. We identified 214 putative *trans*-acting splicing regulators, among which *ZmGRP1*, encoding an hnRNP-like glycine-rich RNA binding protein, regulates the largest *trans*-cluster. Knockout of *ZmGRP1* by CRISPR/Cas9 altered splicing of numerous downstream genes. We found that 739 sQTLs colocalized with previous marker-trait associations, most of which occurred without changes in overall mRNA level. Our findings uncover the importance of AS in diversifying gene function and regulating phenotypic variation.**

## INTRODUCTION

Alternative splicing (AS) of precursor mRNAs (pre-mRNAs) is an essential regulatory mechanism that greatly increases transcriptome and proteome diversity by producing multiple mRNA isoforms from a single gene. Extensive studies have shown that AS is prevalent in various eukaryotic organisms and plays important roles in diverse biological processes (Reddy et al., 2013). In humans, ~95% of genes are subject to AS, with exon-skipping events as the predominant AS type (Wang et al., 2008). It has been estimated that ~15% of genetic diseases are due to mutations affecting splicing (Kornblihtt et al., 2013). In plants, 33 to 60% of intron-containing genes undergo AS (Zhang et al., 2010;

Marquez et al., 2012; Shen et al., 2014; Thatcher et al., 2014; Mandadi and Scholthof, 2015). In contrast to humans, intron retention (IR) is the most prevalent AS event in plants (Reddy et al., 2013). AS functions in a wide range of plant growth and development processes, including flowering time induction, circadian clock control, and plant responses to environmental stress (Staiger and Brown, 2013).

AS arises due to differential usage of alternative splice sites during pre-mRNA splicing. The decision on which splice sites are selected is determined by *cis*-regulatory elements in pre-mRNA and the recognition of these sites by *trans*-acting splicing factors (Reddy et al., 2013; Staiger and Brown, 2013). The *cis*-regulatory sequences include splice sites, polypyrimidine tract, branch point sequences, and exonic and intronic splicing enhancer and silencer sequences, which are binding sites for splicing factors (Perteau et al., 2007; Staiger and Brown, 2013). Serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) are the main families of splicing factors that control splice site choice and guide the spliceosome assembly (Syed et al., 2012; Erkelenz et al., 2013). In general, SR proteins promote intron removal, whereas hnRNPs inhibit splice

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Address correspondence to ft55@cau.edu.cn or yxiaohong@cau.edu.cn. The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Feng Tian (ft55@cau.edu.cn).

<sup>[OPEN]</sup>Articles can be viewed without a subscription.  
www.plantcell.org/cgi/doi/10.1105/tpc.18.00109

## IN A NUTSHELL

**Background:** Alternative splicing (AS) of precursor mRNAs is an essential regulatory mechanism that greatly enhances transcriptome and proteome diversity. AS is prevalent in plants and functions in a wide range of growth and development processes.

**Question:** How is the natural variation in AS regulated? How is the regulation of AS coupled with the gene regulatory network? Is the natural variation in AS and in overall mRNA level commonly or independently regulated? To what extent AS does contribute to phenotypic variation?

**Findings:** We performed a genome-wide association study to identify splicing quantitative trait loci (sQTLs) in developing kernels from 368 diverse maize inbred lines. We identified 19,554 unique sQTLs controlling the splicing of 6570 genes. An average of three sQTLs was mapped for each gene, indicating that the splicing variation of most genes is under relatively simple genetic control. Most sQTLs showed small isoform usage changes without involving major isoform switches between genotypes. Characterization of the regulatory features of sQTLs demonstrated that nonsense-mediated mRNA decay, microRNA-mediated regulation, and small interfering peptide-mediated peptide interference are frequently involved in sQTL regulation, highlighting the importance of AS in coupling different layers of gene regulation mechanisms. We found that the natural variation in AS and overall mRNA level appears to be independently regulated with different *cis*-sequences preferentially used. We identified 214 putative *trans*-acting splicing regulators and further demonstrated that ZmGRP1, an hnRNP-like glycine-rich RNA-binding protein that regulates the largest *trans*-cluster, altered splicing of numerous downstream genes. We found that 739 sQTLs colocalized with previous marker-trait associations, most of which occurred without changes in overall mRNA level, underscoring the importance of AS in regulating phenotypic variation.

**Next steps:** Further work is required to specifically elucidate the molecular mechanisms by which sQTLs function in the gene regulation network and affect the phenotypic variation in the traits of interest.

site selection in the regulation of AS (Syed et al., 2012; Erkelenz et al., 2013; Geuens et al., 2016). The abundance and activity of splicing factors determine the AS profiles of downstream target genes (Syed et al., 2012; Erkelenz et al., 2013).

AS alters the sequence of an RNA transcript, which has important consequences for the cell (Reddy et al., 2013). The main consequence of AS is to increase proteome diversity by producing protein isoforms that differ in subcellular localization, stability, or function (Syed et al., 2012; Reddy et al., 2013). Recent studies have shown that the AS of some transcription factors can generate truncated proteins or polypeptides that act as small interfering peptides (siPEPs) to negatively regulate authentic proteins via peptide interference (PEPI) (Seo et al., 2011b, 2012; Li et al., 2012). However, not all AS events generate functional proteins. AS frequently produces transcript isoforms carrying in-frame premature termination codons (PTCs) targeted for degradation by the nonsense-mediated mRNA decay (NMD) pathway (Kalyna et al., 2012; Drechsel et al., 2013). AS-coupled NMD is a widely conserved eukaryotic pathway for the regulation of transcript levels (Chang et al., 2007; Kurihara et al., 2009; Brogna et al., 2016). In *Arabidopsis thaliana*, ~11 to 18% of splice variants are coupled with NMD (Kalyna et al., 2012; Drechsel et al., 2013). Moreover, microRNAs (miRNAs), a class of endogenous small noncoding RNAs ~20 to 24 nucleotides in length, can regulate complementary mRNAs by inducing translational repression and mRNA decay (Iwakawa and Tomari, 2015). AS of some genes can produce splice variants that either gain or lose binding sites for miRNA (Reddy et al., 2013).

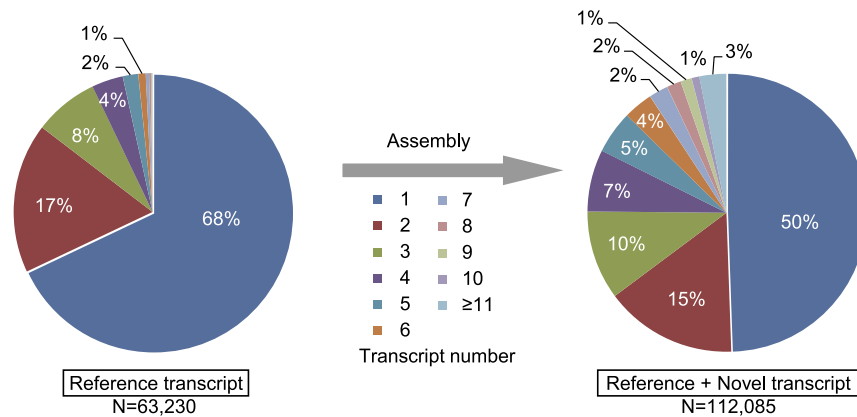
Population-level transcriptome analyses have identified wide natural variation in AS (Wang et al., 2008; Gan et al., 2011; Huang et al., 2015). However, how the natural variation in AS is regulated and coupled with the gene regulation network and contribute to phenotypic variation is poorly understood. Recently, several large-scale genome-wide association studies have been

conducted to identify splicing quantitative trait loci (sQTLs) controlling AS variation in human populations (Pickrell et al., 2010b; Monlong et al., 2014; GTEx Consortium, 2015; Zhang et al., 2015; Li et al., 2016; Takata et al., 2017). These analyses demonstrated that sQTLs can explain a substantial proportion of marker-trait associations. By contrast, few sQTL studies have been conducted in plants. Thatcher et al. (2014) and Mei et al. (2017) analyzed the AS differences between two maize (*Zea mays*) inbred lines, B73 and Mo17, in a linkage population. To better understand the genetic regulatory mechanism of AS and evaluate the relative contribution of AS to trait variation at the species level, large-scale AS analysis in a wide range of maize genotypes is needed. In this study, we performed a genome-wide association study (GWAS) to identify sQTLs in developing kernels from 368 diverse maize inbred lines. Characterization of the regulatory features of sQTLs demonstrated that AS plays a crucial role in coupling different layers of gene regulation network and regulating phenotypic variation in maize.

## RESULTS

### Transcriptome Assembly and sQTL Mapping

Using RNA sequencing, Fu et al. (2013) profiled the transcriptome of the developing kernels at 15 d after pollination (DAP) from 368 maize inbred lines. To characterize the AS variation in the population, we performed a population-level transcript assembly based on the B73 reference genome (AGPv3) in this study. We discovered 48,855 potential novel transcripts from known genes. These results showed that the percentage of protein-coding genes that undergo AS increased from 32% (12,616 genes) to 50% (19,844 genes) (Figure 1), indicating that maize



**Figure 1.** Transcriptome Assembly.

The distribution of transcript isoform number before and after transcriptome assembly. The proportion of multi-isoform genes increased from 32 to 50%.

undergoes AS at levels similar to other model plants (Marquez et al., 2012; Mandadi and Scholthof, 2015).

A total of 12,150 expressed genes that have at least two isoforms were used for sQTL mapping. For each expressed gene, the relative ratio of total gene expression represented by each isoform (designated as the splicing ratio) was calculated in each inbred line and used as the splicing phenotype in sQTL mapping. By combining 1.25 million single-nucleotide polymorphisms (SNPs) with minor allele frequency  $\geq 0.05$  (Liu et al., 2017), GWAS was performed to identify sQTLs using a mixed linear model that corrects for population structure, unknown confounders, and family relatedness (Yu et al., 2006). At a conservative Bonferroni-corrected  $P < 7.43 \times 10^{-7}$ , we identified 222,481 SNPs showing significant associations with splicing variation in 7869 genes. From the initial associations, we performed two steps of filtering. First, for each gene detected with sQTL, we conducted linkage disequilibrium (LD) analysis for the associated SNPs to identify independently associated SNPs. A unique sQTL was defined when the associated SNP was not in LD ( $r^2 < 0.1$ ) with any other associated SNPs on the same chromosome for the target gene. Second, we examined the splicing ratio changes at each unique sQTL. Only sQTLs that showed more than 5% difference in splicing ratio between genotypes were retained for the subsequent analysis. We finally obtained 19,554 unique sQTLs regulating the AS of 6570 genes (representing 54% of expressed genes for sQTL analysis) (Figure 2A; Supplemental Data Set 1). To verify the sQTLs, we randomly selected 20 genes detected with a sQTL and performed RT-PCR analysis to examine the isoform relative expression differences at the sQTL. Of the 20 examined genes, 17 genes (85%) exhibited consistent isoform relative expression differences between genotypes with the splicing ratio differences detected in sQTL analysis (Supplemental Figure 1), suggesting that the sQTLs detected in this study are highly reproducible.

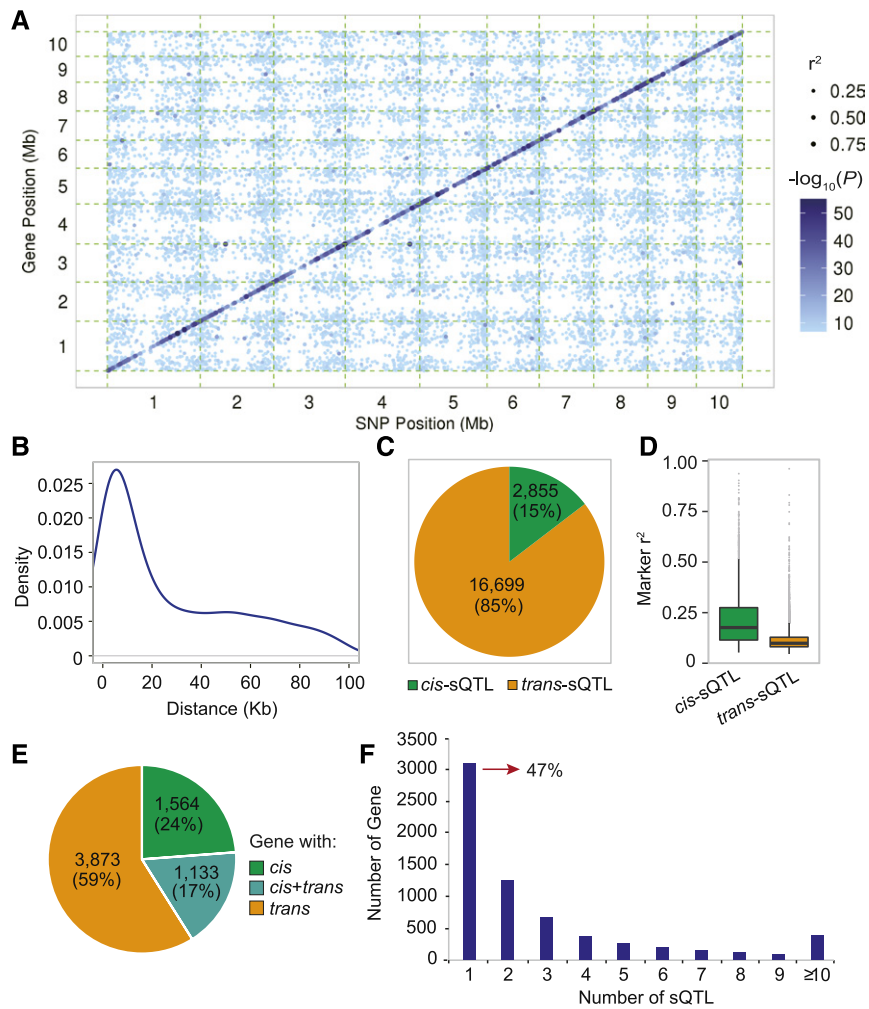
We analyzed the relative positions of sQTLs and their associated genes (Figure 2B), which showed a similar distribution as that for the relative positions of eQTLs and their associated

genes (Fu et al., 2013). An obvious inflection point occurred at the position of 20 kb (Figure 2B). Therefore, a sQTL was considered a *cis*-sQTL when the associated SNP was within 20 kb of the transcription start site or transcription end site of the target gene; otherwise, the sQTL was considered *trans*-sQTL. Among the mapped sQTLs, 2855 (15%) and 16,699 (85%) loci were *cis*- and *trans*-sQTLs, respectively (Figure 2C; Supplemental Data Set 1). *Cis*-sQTLs explained more splicing variation than *trans*-sQTLs (Figure 2D). Among the genes detected with sQTLs, 41% were mapped with *cis*-sQTLs (Figure 2E). An average of three sQTLs was mapped for each gene, with 47% of mapped genes having only a single sQTL affecting their splicing variation (Figure 2F), suggesting that the splicing variation of most genes is under relatively simple genetic control.

To analyze the functional features of genes detected with sQTLs, we conducted functional annotation for the 6570 genes detected with sQTLs and analyzed the distribution of these genes across functional categories. Pathway and Gene Ontology (GO) term enrichment analysis showed that genes detected with sQTLs are significantly overrepresented in a number of pathways and GO categories, among which the most significant categories include protein metabolism, RNA metabolism, response to stimulus, and cellular process (Supplemental Data Sets 2 and 3). These results indicated that genes with sQTLs function in diverse biological processes.

### sQTL-Associated AS Effects

By comparing the splice junctions of the isoform pairs differentially used at sQTLs, we analyzed the AS types that each sQTL involved. In total, 20,317 sQTL-associated AS events were identified, with 81% of sQTLs associated with one event (Supplemental Figure 2). Among these AS events, IR was the most frequent (36%), followed by alternative acceptor site (AA; 22%), exon skipping (ES; 14%), and alternative donor site (AD; 13%) (Figure 3A).



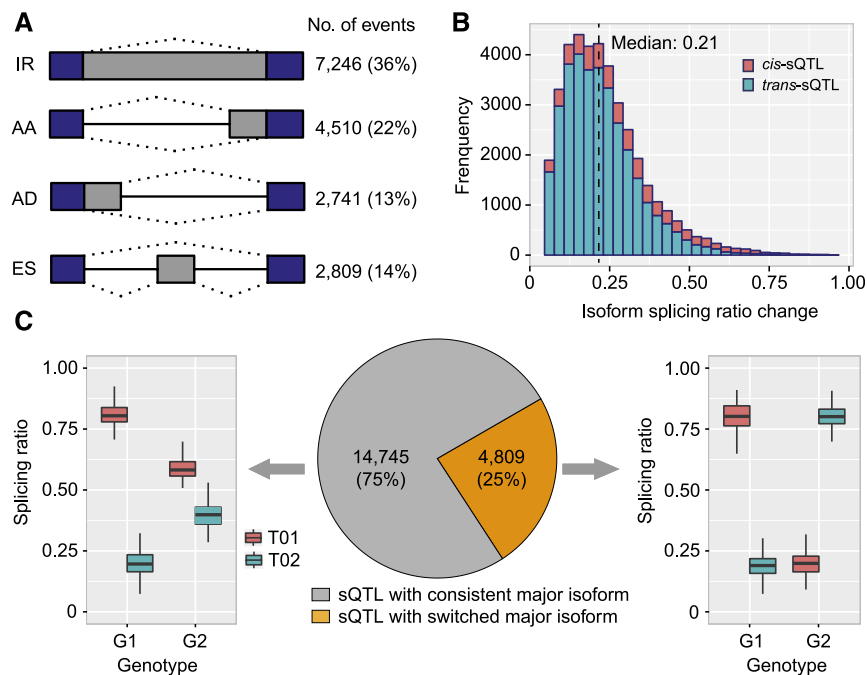
**Figure 2.** sQTL Mapping.

(A) sQTL heat map. The x axis and y axis indicate the physical position of the lead SNP of the sQTL and its associated gene, respectively. Each dot represents a sQTL. Splicing variation explained by sQTLs ( $r^2$ ) is shown in gradual dot size, and the P value of sQTLs is shown in gradual colors.  
 (B) The density of distance between sQTLs and their regulated gene.  
 (C) The distribution of *cis*- and *trans*-sQTLs.  
 (D) The distribution of splicing variation explained by sQTLs ( $r^2$ ). *Cis*-sQTLs explained more splicing variation than *trans*-sQTLs.  
 (E) The distribution of genes regulated by *cis*- and/or *trans*-sQTLs.  
 (F) Number of sQTLs identified for each gene. The x axis and y axis represent the number of sQTLs identified for each gene and the number of genes in each group, respectively. “47%” indicates the proportion of genes detected with a single sQTL.

To analyze the effect of sQTLs on the isoform relative expression pattern, we computed the splicing ratio difference between genotypes at each sQTL. As shown in Figure 3B, the distribution of splicing ratio changes was left skewed, with a median splicing ratio change of 21%. *Cis*-sQTLs and *trans*-sQTLs showed a similar effect on the relative changes in isoform expression (Figure 3B). Furthermore, we calculated the number of sQTLs showing a major isoform switch between genotypes. Interestingly, 25% of sQTLs were associated with a switch in major isoform between genotypes (Figure 3C). The majority of sQTLs (75%) showed small relative splicing ratio changes without involving major isoform switches between genotypes (Figure 3C).

**sQTL-Affected Isoforms Tend to Display Distinct Protein Functions**

A total of 14,596 transcript isoforms involving 6570 genes were affected by sQTLs. We assessed the extent to which the isoform pairs differentially used at sQTLs are functionally similar or different from each other at the protein level. To address this question, we predicted the protein domains of sQTL-affected isoforms using HMMER3 (Eddy, 2011) and compared the difference in protein domain configuration for each isoform pair. Out of 6570 genes detected with sQTL, the sQTL of 1977 genes involved differential use of isoform pairs that gained, lost, or



**Figure 3.** sQTL-Associated AS Effects.

**(A)** Categorization of sQTL-associated AS events. IR, intron retention; AA, alternative acceptor; AD, alternative donor; ES, exon skipping.

**(B)** The distribution of splicing ratio differences between genotypes at each sQTL.

**(C)** The distribution of isoform usage pattern. The left pattern indicates that sQTLs only affect the relative usage ratio of isoforms without involving major isoform switch between genotypes. The right pattern indicates that sQTLs involve major isoform switch with different major isoforms used between genotypes.

exchanged domains relative to each other (Supplemental Data Set 4), which is unlikely to have occurred by chance alone ( $P < 0.001$ , 1000 permutations) (Supplemental Figure 3A). This result suggested that sQTLs tend to regulate isoforms with different protein functions. However, we should note that the prediction of the fate of transcript isoforms we performed here is based on computational analysis; whether they are actually translated in the predicted manner requires systematic experimental testing.

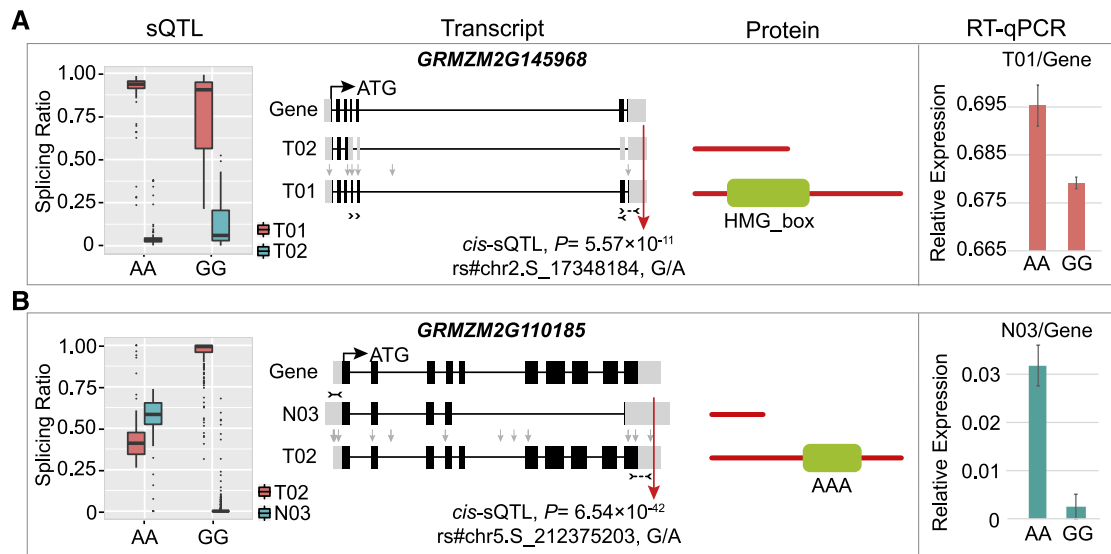
For example, *GRMZM2G145968* was detected with a significant *cis*-sQTL (rs#chr2.S\_17348184,  $P = 5.57 \times 10^{-11}$ ) that was associated with differential use of two isoforms that differ by an HMG\_box domain that is present in isoform T01 but absent in isoform T02 (Figure 4A). *GRMZM2G110185* was detected with a significant *cis*-sQTL (rs#chr5.S\_212375203,  $P = 6.54 \times 10^{-42}$ ) that was associated with differential use of two isoforms that differ by an AAA domain that is present in isoform T02 but absent in isoform N03 (Figure 4B). The differential uses of isoforms at these sQTLs were validated by RT-qPCR assay (Figures 4A and 4B).

### sQTL-Affected Isoforms Act as siPEPs

Recent studies have shown that the AS of some transcription factors could generate truncated splice variants possessing dimerization domains but lacking the functional domains required for DNA binding and/or transcriptional regulation (Seo

et al., 2011a, 2013; Staudt and Wenkel, 2011). These truncated proteins can act as dominant-negative siPEPs to competitively inhibit the corresponding full-size transcription factor activity by forming nonfunctional heterodimers (Seo et al., 2011a, 2013; Staudt and Wenkel, 2011). This protein-level regulation has been designated PEPI, which is conceptually similar to the RNA interference mediated by small interfering RNAs (siRNAs) (Seo et al., 2011a, 2013). To examine whether siPEP-mediated PEPI is frequently employed by sQTLs to regulate genotypic AS variation, we performed protein structural analysis for the transcripts of transcription factors detected with sQTLs. Among the 6570 genes detected with sQTLs, 772 genes encoded transcription factors, at which a total of 2148 sQTLs were mapped. Within each transcription factor detected with a sQTL, we compared the protein domain organizations of transcript isoforms differentially used at the target sQTL. Interestingly, the sQTLs of 161 transcription factors involved differential use of the full-size transcription factor and the truncated splice isoform lacking functional DNA binding domains, which is unexpected by chance alone ( $P < 0.001$ , 1000 permutations) (Supplemental Figure 3B). This result suggested that PEPI is frequently involved in the sQTL regulation of transcription factors. These truncated isoforms, if present, could act as competitive siPEPs in regulating the activity of target transcription factors.

For example, *GRMZM2G169382*, a gene encoding an AP2 transcription factor, was detected with a significant *trans*-sQTL



**Figure 4.** Examples of sQTL-Affected Isoforms Displaying Different Protein Domains.

Box plots display splicing ratio change at different genotypes of sQTL SNPs. For the transcript model, UTRs are shown in gray boxes on the two sides, and CDSs are shown in black boxes in the middle. Lines indicate introns. Transcripts named “T” and “N” stand for known and newly assembled transcript, respectively. Red arrow indicates the position of *cis*-sQTL SNP, and gray arrows indicate the position of other less significant SNPs. The positions of primers used for RT-qPCR were indicated by dotted lines with arrows. For the protein model, the conserved domain between different isoforms is indicated in green. The RT-qPCR results are shown on the right. For each genotype at the sQTL, four inbred lines were randomly selected from the maize association panel for the RT-qPCR assay of kernels at 15 DAP. Values represent the mean  $\pm$  sd of the measurements ( $n = 4$ ).

**(A)** Significant *cis*-sQTL detected at *GRMZM2G145968* that involved differential use of two isoforms differing in an HMG\_box domain. AA and GG represent the two genotype groups at sQTL SNP rs#chr2.S\_17348184.

**(B)** Significant *cis*-sQTL detected at *GRMZM2G110185* that involved differential use of two isoforms differing by an AAA domain. AA and GG represent the two genotype groups at sQTL SNP rs#chr5.S\_212375203.

(rs#chr3.S\_56287892,  $P = 3.69 \times 10^{-24}$ ) at which two isoforms (T01 and T02) showed significant splicing ratio differences between genotypes (Figures 5A and 5B). Interestingly, the isoform T01 contained an AP2 DNA binding domain, which would have functional transcription factor activity, while isoform T02 lost the DNA binding domain (Figure 5C). The differential use of isoforms at the sQTL was validated by RT-qPCR assays (Figure 5D). The *trans*-sQTL of *GRMZM2G169382* is likely mediated by a PEP1 mechanism.

### AS-Coupled NMD Mediates sQTL Regulation

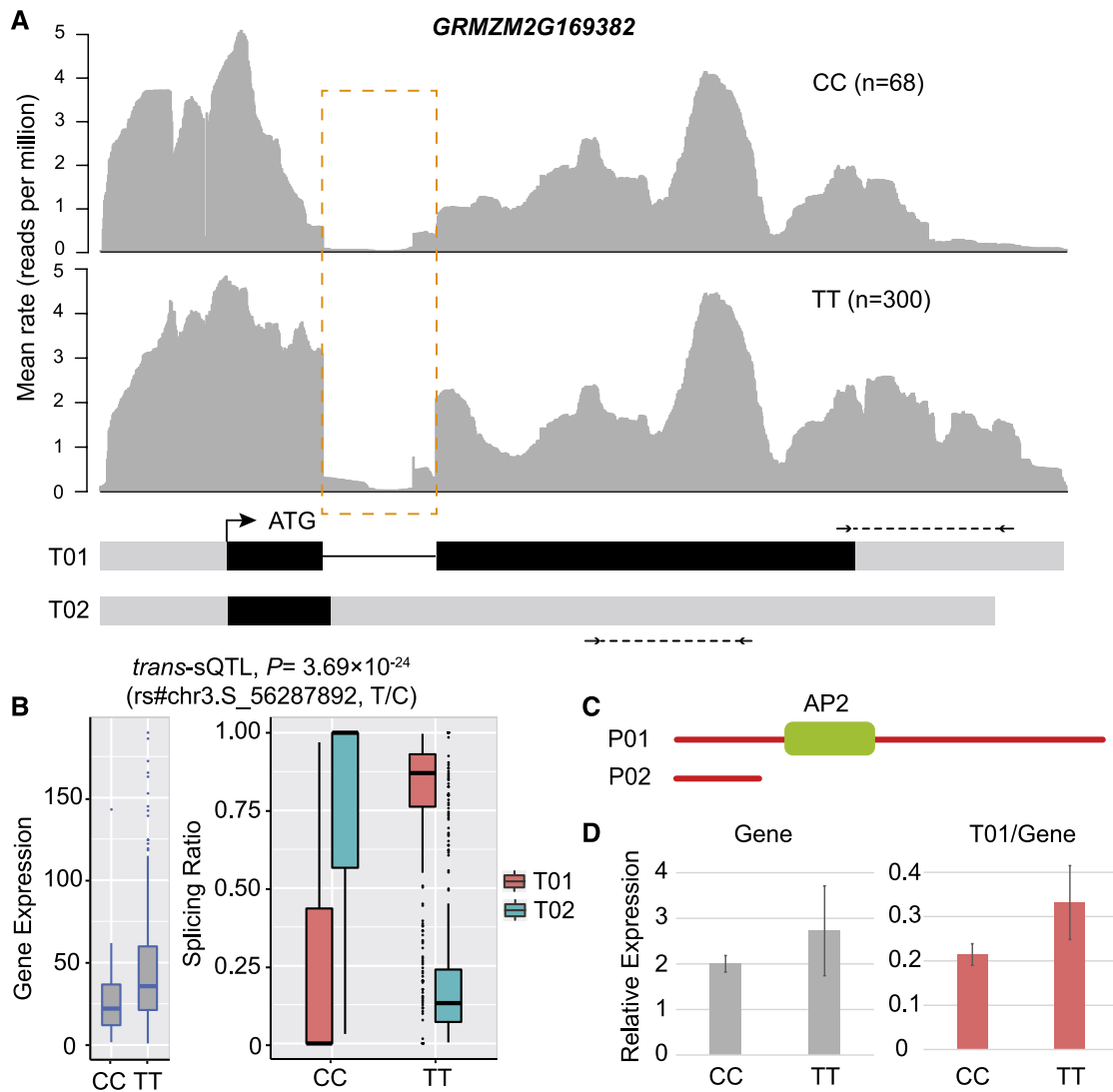
Alternative splicing of many genes can generate isoforms that contain a premature stop codon (PTC) that can be recognized by NMD degradation pathway, thereby regulating the level of functional transcripts (Chang et al., 2007; Brogna et al., 2016). However, unlike in mammals, transcripts with retained introns in plants are often insensitive to NMD owing to their retention in the nucleus (Kim et al., 2009; Kalyna et al., 2012; Göhring et al., 2014). To assess whether NMD plays an important role in mediating sQTL regulation, we analyzed whether sQTL-affected transcript isoforms are potential targets of NMD. Transcript isoforms without IR events carrying a PTC located at least 50 nucleotides upstream of an exon splice junction, 3'-untranslated regions (UTRs) longer than 350 bp, introns within 3'-UTRs, and upstream open reading

frames greater than 35 amino acids within 5'-UTRs (Kalyna et al., 2012; Drechsel et al., 2013) were marked as potential NMD candidates. Interestingly, the sQTLs of 1643 genes involved the differential use of isoform pairs that gain or lost NMD features (Supplemental Data Set 5), which is unlikely to have occurred by chance alone ( $P < 0.001$ , 1000 permutations) (Supplemental Figure 3C). This result suggested that NMD is frequently coupled with AS in modulating genotypic splicing variation.

For instance, *GRMZM2G021742*, a gene encoding U2 small nuclear ribonucleoprotein A, was detected with a significant *cis*-sQTL (rs# chr1.S\_298434582,  $P = 4.8 \times 10^{-12}$ ), where the two isoforms T01 and T02 were differentially used between genotypes (Figure 6A). No significant difference in overall mRNA abundance was detected at this *cis*-sQTL (Figure 6B). The alternative acceptor sites of the 5th exon resulted in T02 harboring a PTC located 79 bp upstream of the exon-exon junction, and T02 showed low expression levels relative to T01, suggesting that T02 might be targeted and degraded by NMD. The relative isoform expression difference of T02 was validated by RT-qPCR (Figure 6C).

### A miRNA-Mediated Mechanism Is Involved in sQTL Regulation

miRNAs are a class of small noncoding RNAs that typically bind to complementary sequences in mRNA targets and attenuate



**Figure 5.** AS-Coupled PEPi.

**(A)** An example of the AS-coupled PEPi mechanism. *GRMZM2G169382* was detected with a significant *trans*-sQTL. For the transcript model, UTRs are shown in gray boxes on the two sides, and CDSs are shown in black boxes in the middle. Lines indicate introns. Transcripts named “T” stand for known transcript. The positions of primers used for RT-qPCR were indicated by dotted lines with arrows. For the upper panel, the average rate at each base on the gene is plotted, where individuals were stratified according to their genotype at sQTL SNP rs#chr3.S\_56287892. Brown dotted box shows where transcript structures differ.

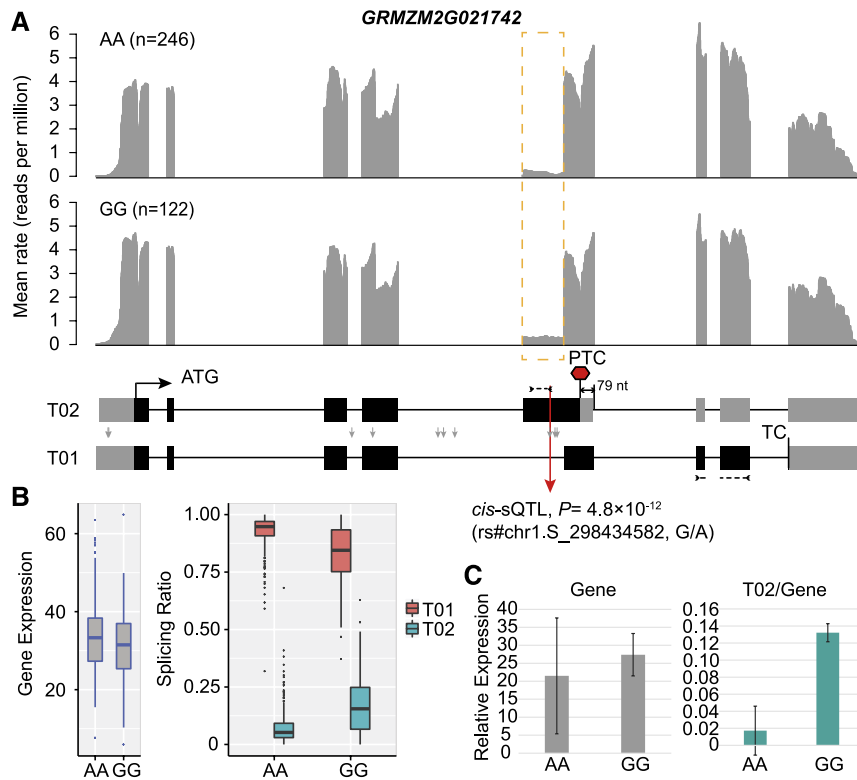
**(B)** The box plot shows the difference in splicing ratio and overall gene expression level between genotypes (CC and TT) at sQTL SNP rs#chr3.S\_56287892.

**(C)** Structure of P02, showing it has lost the whole DNA binding domain relative to P01.

**(D)** RT-qPCR results. For each genotype at the sQTL, four inbred lines were randomly selected from the maize association panel for the RT-qPCR assay of kernels at 15 DAP. Values represent the mean  $\pm$  SD of the measurements ( $n = 4$ ).

gene expression (Iwakawa and Tomari, 2015). AS of some genes can generate transcript isoforms that either contain or lack binding sites for miRNA. To assess whether a miRNA-mediated mechanism is involved in sQTL regulation, we performed bioinformatic analysis to examine whether sQTL-affected transcript isoforms are potential targets of miRNAs. Bioinformatics analysis showed that there are 334 genes that are potential targets of 70 high-confidence mature miRNAs. Of these 334 miRNA target

genes, 63 genes were detected with significant sQTLs. Among them, 16 genes involved differential use of isoform pairs that gained or lost miRNA target sites relative to each other (Supplemental Data Set 6), which is unexpected by chance alone ( $P = 0.026$ , 1000 permutations) (Supplemental Figure 3D). We should note that this is a conservative analysis because current miRNA annotation in maize is limited. sQTLs might be prevalently coupled with miRNA regulation.



**Figure 6.** AS-Coupled NMD.

**(A)** An example of an AS-coupled NMD mechanism. *GRMZM2G021742* was detected with a significant *cis*-sQTL. For the transcript model, UTRs are shown in gray boxes on the two sides, and CDSs are shown in black boxes in the middle. Lines indicate introns. Transcripts named “T” stand for known transcript. Red arrow indicates the position of *cis*-sQTL SNP, and gray arrows indicate the position of other less significant SNPs. TC, termination codon; PTC, premature termination codon. The positions of primers used for RT-qPCR were indicated by dotted lines with arrows. For the upper panel, the average rate at each base on the gene is plotted, where individuals were stratified according to their genotype (AA and GG) at sQTL SNP rs#chr1.S\_298434582. Brown dotted box shows where transcript structures differ. Relative to T01, T02 has a PTC located 79 bp upstream of the exon-exon junction.

**(B)** Significant changes in splicing ratio at sQTL SNP rs#chr1.S\_298434582 but no difference in overall gene expression level.

**(C)** RT-qPCR results. For each genotype at the sQTL, four inbred lines were randomly selected from the maize association panel for the RT-qPCR assay of kernels at 15 DAP. Values represent the mean  $\pm$  SD of the measurements ( $n = 4$ ).

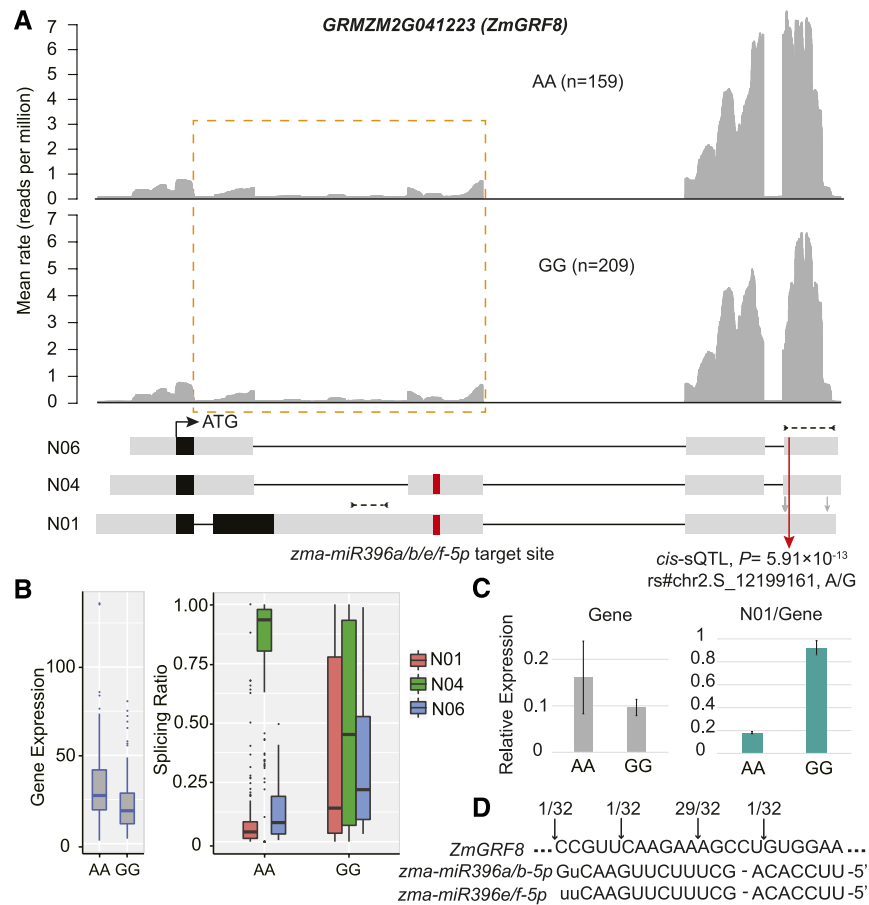
For example, *GRMZM2G041223* (*ZmGRF8*), encoding GRF (GROWTH-REGULATING FACTOR)-transcription factor 8, was detected with a significant *cis*-sQTL (rs#chr2.S\_12199161,  $P = 5.91 \times 10^{-13}$ ), where the isoforms N01, N04, and N06 were differentially used between genotypes (Figures 7A and 7B). The relative isoform expression difference between transcripts was validated by RT-qPCR (Figure 7C). Interestingly, N01 and N04 were predicted to contain miRNA binding sites for *zma-miR396a/b/e/f-5p*, whereas N06 lost the miRNA binding sites (Figure 7A). To validate the miR396 targeting *ZmGRF8*, RLM-5'-RACE assays were used to identify the cleavage sites. Our data showed that miR396 led to *ZmGRF8* RNA cleavage in developing maize kernels (Figure 7D).

**AS and Overall mRNA Level Are under Relatively Independent Genetic Control**

AS is primarily a cotranscriptional process (Reddy et al., 2013; Nafteberg et al., 2015). The factors affecting AS may also

influence overall gene expression levels. However, the extent to which the natural variation in overall mRNA levels and AS is commonly regulated remains poorly understood. We assessed the overlap of sQTLs mapped in this study with eQTLs previously identified in the same set of samples (Liu et al., 2017). After filtering for genes used to map sQTLs in this study, a total of 7696 previously identified *cis*-eQTLs were used for comparison with *cis*-sQTLs. Only those *cis*-sQTLs and *cis*-eQTLs that were located within 100 kb and in LD ( $r^2 \geq 0.1$ ) and regulate the same gene are considered as overlapped loci. Interestingly, we found that only 1504 SNPs (16.6%) were simultaneously associated with changes in total mRNA level and AS of the same gene (Figure 8A), indicating that the natural variation in total mRNA levels and AS tend to be under relatively independent genetic control. Examples shown in Figure 8 are representative genes for each regulation mode. A significant *cis*-sQTL was detected at *su2* (*sugary2*; *GRMZM2G348551*), which encodes a starch-branching enzyme, but this *cis*-sQTL had no significant





**Figure 7.** AS-Coupled miRNA Regulation.

**(A)** An example of the AS-coupled miRNA mechanism. *GRMZM2G041223* was detected with a significant *cis*-sQTL. For the transcript model, UTRs are shown in gray boxes on the two sides, and CDSs are shown in black boxes in the middle. Lines indicate introns. Transcripts named “N” stand for newly assembled transcript. Red arrow indicates the position of *cis*-sQTL SNP, and gray arrows indicate the position of other less significant SNPs. The positions of primers used for RT-qPCR are indicated by dotted lines with arrows. For the upper panel, the average rate at each base on the gene is plotted, where individuals were stratified according to their genotype (AA and GG) at sQTL SNP rs#chr2.S\_12199161. Brown dotted box shows where transcript structures differ. Red box indicates the position of miRNA target site.

**(B)** Significant changes in splicing ratio and overall gene expression level at sQTL SNP rs#chr2.S\_12199161.

**(C)** RT-qPCR results. For each genotype at the sQTL, four inbred lines were randomly selected from the maize association panel for the RT-qPCR assay of kernels at 15 DAP. Values represent the mean  $\pm$  SD of the measurements ( $n = 4$ ).

**(D)** RLM-5'-RACE mapping of *ZmGRF8* mRNA cleavage sites. The numbers of clones with particular termini are indicated by arrows.

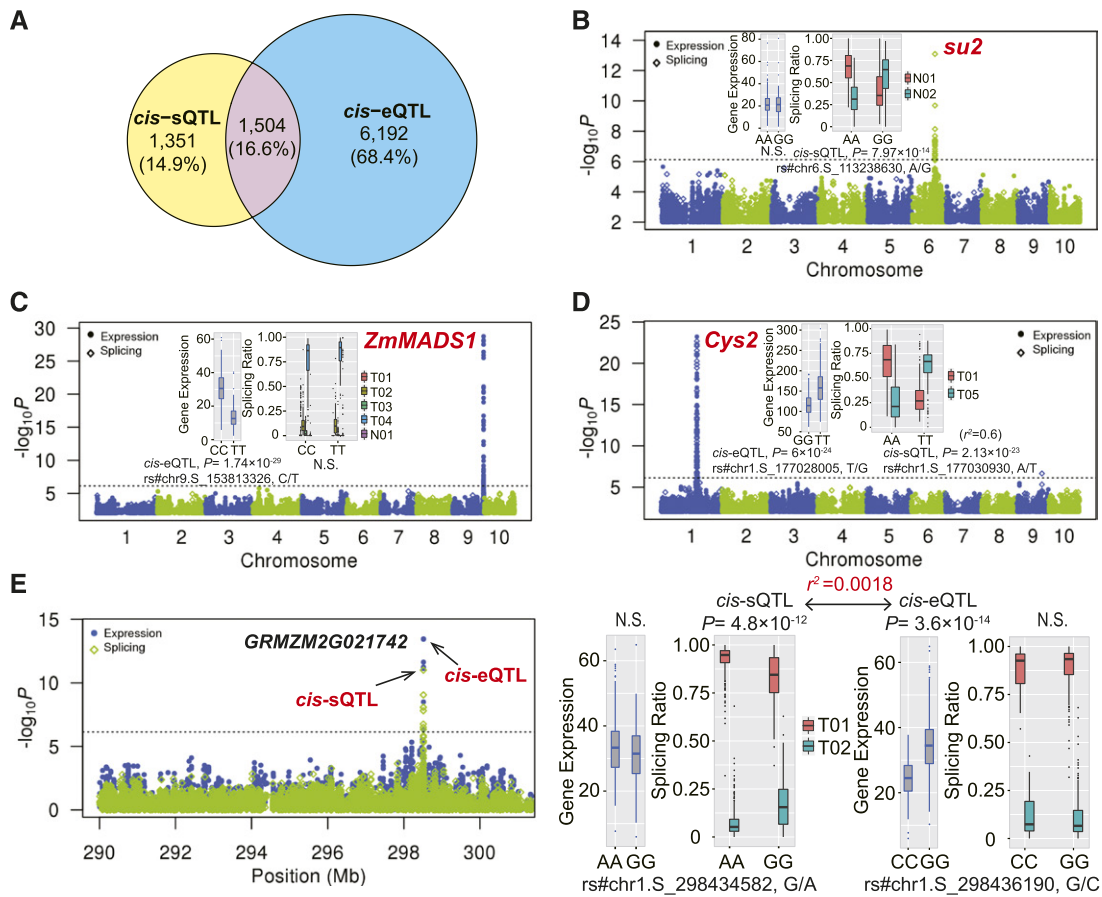
effect on the total gene expression level (Figure 8B). A significant *cis*-eQTL was detected at *ZmMADS1* (*MADS box protein1*; *GRMZM2G171365*), but this *cis*-eQTL had no significant effect on AS (Figure 8C). Different from *su2* and *ZmMADS1*, both *cis*-sQTL and *cis*-eQTL were detected at *Cys2* (*Cysteine synthase2*; *GRMZM2G005887*), and these two SNPs are in strong LD ( $r^2 = 0.6$ ) (Figure 8D), indicating that the underlying variant simultaneously impacted the AS and mRNA level of *Cys2*.

Interestingly, 676 genes were simultaneously detected with *cis*-sQTL and *cis*-eQTL, but the associated SNPs were not in LD ( $r^2 < 0.1$ ), suggesting that the overall mRNA level and AS of these genes are subject to independent *cis*-regulation (Supplemental Data Set 7). For example, both significant *cis*-

eQTL (rs#chr1.S\_298436190,  $P = 3.6 \times 10^{-14}$ ) and *cis*-sQTL (rs#chr1.S\_298434582,  $P = 4.8 \times 10^{-12}$ ) were detected at *GRMZM2G021742*, which encodes U2 small nuclear ribonucleoprotein A (Figure 8E). However, rs#chr1.S\_298436190 and rs#chr1.S\_298434582 that were 1.6 kb apart were not in LD ( $r^2 = 0.0018$ ; Figure 8E), indicating that different *cis*-variants controlled the natural variation in AS and mRNA level at *GRMZM2G021742*.

### *Cis*-sQTLs Are Significantly Enriched in Splice-Site Regions

To examine why the natural variation in AS and mRNA level showed relatively independent genetic control, we analyzed

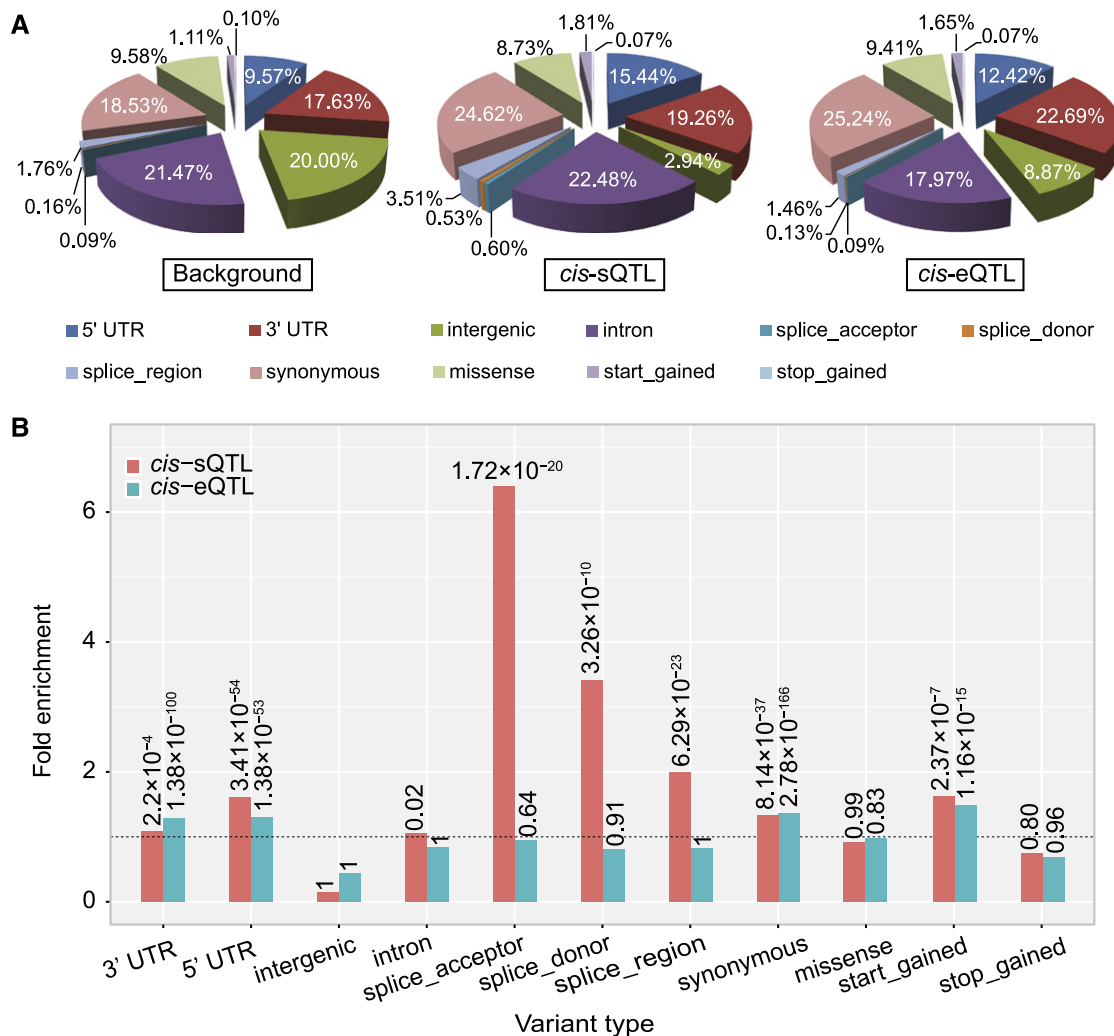


**Figure 8.** Relatively Independent Genetic Control of sQTLs and eQTLs.

(A) Venn diagram showing the overlap between sQTLs and eQTLs. From (B) to (E), Manhattan plots show GWAS results for splicing ratios and overall expression levels for each gene. Transcripts named “T” and “N” stand for known and newly assembled transcript, respectively. (B) An example of a gene (*su2*) with only *cis*-sQTLs detected without affecting the expression level. Box plots display splicing ratio change at different genotypes (AA and GG) of sQTL SNP rs#chr6.S\_113238630. (C) An example of a gene (*ZmMADS1*) with only *cis*-eQTLs detected without affecting the splicing. Box plots display total mRNA level change at different genotypes (CC and TT) of eQTL SNP rs#chr9.S\_153813326. (D) An example of a gene (*Cys2*) with detected variants affecting both AS and total mRNA level. Box plots display splicing ratio change at different genotypes (AA and TT) of sQTL SNP rs#chr1.S\_177030930 and total mRNA level change at different genotypes (GG and TT) of eQTL SNP rs#chr1.S\_177028005. (E) Independent *cis*-regulation of AS and total mRNA level. Both *cis*-sQTLs and *cis*-eQTLs were detected at *GRMZM2G021742*, but the two SNPs were not in LD ( $r^2 = 0.0018$ ). Box plots display splicing ratio change at different genotypes (AA and GG) of sQTL SNP rs#chr1.S\_298434582 and total mRNA level change at different genotypes (CC and GG) of eQTL SNP rs#chr1.S\_298436190.

*cis*-sQTL and *cis*-eQTL associated variant types and their relative distributions (Figure 9A). All 1.25 million SNPs were used as the background SNP set to evaluate the relative enrichment of *cis*-sQTL and *cis*-eQTL SNPs in each variant category. As shown in Figure 9B, *cis*-sQTLs were significantly overrepresented in UTRs, 5' splice donor sites, 3' splice acceptor sites, regions flanking splice sites (within one to three bases of an exon or three to eight bases of an intron), intronic regions, start\_gained, and synonymous variants. Strikingly, we observed more than threefold enriched *cis*-sQTLs at the splice donor/acceptor sites, highlighting their crucial roles in modulating splicing variation. The other enriched regions such as regions flanking splice sites

and intronic regions might contain exonic and intronic splicing enhancer and silencer sequences that have been shown to be important for AS (Perteau et al., 2007; Staiger and Brown, 2013). By contrast, for *cis*-eQTLs, significant relative enrichment was detected only in UTRs, start\_gain, and synonymous variants (Figure 9B). No significant enrichments in splice-site regions were detected for *cis*-eQTLs (Figure 9B). This contrasting distribution of variants for *cis*-sQTLs and *cis*-eQTLs suggested that different *cis*-regulatory sequences are preferentially used in the *cis*-regulation of AS and mRNA levels, which might partially explain the relative independent genetic control of gene expression levels and AS.



**Figure 9.** Characterization of the SNP Effect.

(A) Pie charts indicating proportions of SNPs annotated with each functional category.

(B) Relative enrichment of *cis*-sQTLs and *cis*-eQTLs in each functional category. The x axis shows different types of variants. The y axis indicates the fold enrichment of *cis*-sQTLs and *cis*-eQTLs in each variant relative to the background SNPs. P values were calculated using a hypergeometric test.

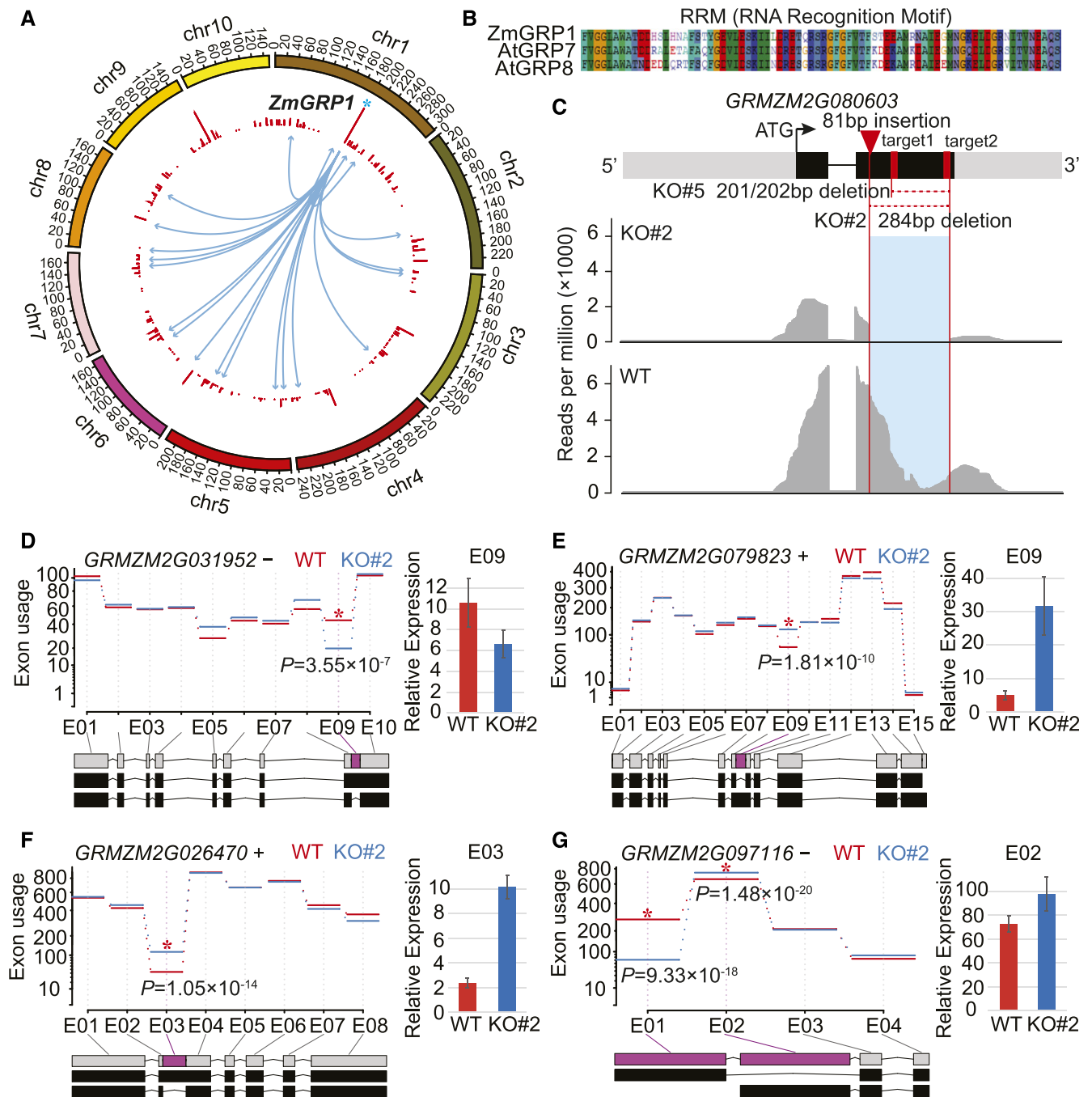
### Trans-Acting Splicing Regulators

In addition to the *cis*-signals in pre-mRNAs, the regulation of AS depends on *trans*-acting splicing regulators that recognize pre-mRNA sequence elements. To identify such *trans*-acting factors, we defined *trans*-clusters as *trans*-sQTL SNPs associated with the splicing variation of more than two *trans*-regulated targets. For each *trans*-cluster, the genes where the most significant *trans*-associated splicing SNPs are located or nearby were considered potential *trans*-acting splicing regulators. After merging *trans*-clusters with their *trans*-associated splicing SNPs in LD ( $r^2 \geq 0.1$ ), a total of 214 putative *trans*-acting splicing regulators were identified (Supplemental Data Set 8). Gene annotations for these putative *trans*-regulators showed that many of them are transcription factors, splicing factors, and splicing-related proteins (Supplemental Data Set 8). Most of these

putative *trans*-regulators (67%) were detected with *cis*-sQTL, *cis*-eQTL, or both (Supplemental Data Set 8).

### ZmGRP1 Functions as a Splicing Factor

Among the putative *trans*-regulators, *ZmGRP1* regulates the largest *trans*-cluster that affects the AS of 21 downstream genes (Figure 10A; Supplemental Data Set 8). *ZmGRP1* is an hnRNP-like glycine-rich RNA binding protein that shows 75% and 74% protein sequence identity with AtGRP7 and AtGRP8 in Arabidopsis (Figure 10B), which have been shown to play important roles in pre-mRNA splicing (Streitner et al., 2010, 2012). To further validate the function of *ZmGRP1*, we knocked out endogenous *ZmGRP1* using CRISPR/Cas9 technology (Belhaj et al., 2015). Two 20-bp sequences in the second exon of *ZmGRP1* were selected as target sites for Cas9 cleavage (Figure 10C).



**Figure 10.** *ZmGRP1* Functions as a *Trans*-Acting Splicing Factor.

(A) *Trans*-acting splicing regulators. The histogram shows the distribution of *trans*-clusters. The lines with arrows indicate that *ZmGRP1* regulates the AS of 21 downstream genes.

(B) The protein sequence alignment of RRM domain of *ZmGRP1* and its Arabidopsis homologs, *AtGRP7* and *AtGRP8*.

(C) The homozygous mutation line of CRISPR/Cas9, with an 81-bp insertion and a 284-bp deletion in the coding region that truncated the *ZmGRP1* ORF. The upper panel shows the gene structure of *ZmGRP1*. Exons and introns were shown as boxes and lines, respectively. CDSs are shown in black and the two target sites are shown in red boxes. The lower panel represents RNA-seq read alignment of *ZmGRP1* for both KO#2 and the wild type.

(D) to (G) Examples of genes with differential exon usages. The results of differential exon usages from the mRNA sequencing are shown on the left, and the corresponding RT-qPCR results are shown on the right. Immature kernels of KO#2 and wild-type plants were collected at 15 DAP for mRNA sequencing and RT-qPCR analysis. Values represent the mean  $\pm$  SD of the measurements for wild type ( $n = 4$  biological replications) and KO#2 ( $n = 3$  biological replications). Exons in red that exhibited significantly differential usage between KO#2 and wild-type plants were indicated by asterisk.

Two homozygous mutation lines (KO#2 and KO#5) were generated with large insertion and deletion in the coding region that truncated the *ZmGRP1* open reading frame (ORF; Figure 10C; Supplemental Figure 4). To examine the effect of the loss of function of *ZmGRP1* on downstream genes, we sequenced mRNAs of immature kernels of KO#2 and wild-type plants collected at 15 DAP. Exon usage analysis showed that a total of 1039 genes ( $q < 0.05$ ) exhibited significant exon usage differences between KO#2 and wild-type plants (Supplemental Data Set 9). To validate the mRNA sequencing results, we randomly selected seven genes with differential exon usages for RT-qPCR analysis (Figures 10D to 10G; Supplemental Figure 5). All seven genes showed consistent exon expression differences with the mRNA sequencing results. GO analysis showed that the genes detected with significant exon usage differences were enriched in small molecule metabolic process, mRNA processing, and translation initiation (Supplemental Data Set 10).

### AS Contributes to Trait Variation

To evaluate the importance of AS for phenotypic variation, we determined whether the sQTLs identified in this study were mapped to the 2351 marker-trait associations previously identified in the same population, including QTLs for metabolite levels (Wen et al., 2014; Deng et al., 2017) and phenotypic traits of kernels (Li et al., 2013; Yang et al., 2014; Liu et al., 2017). A total of 739 unique sQTLs were colocalized ( $<100$  kb;  $r^2 \geq 0.1$ ) with previously identified trait QTLs (Figure 11A; Supplemental Data Set 11). Notably, 611 sQTLs (83%) were associated with trait variation in the absence of changes in overall gene expression levels, strongly indicating the importance of AS in regulating phenotypic variation.

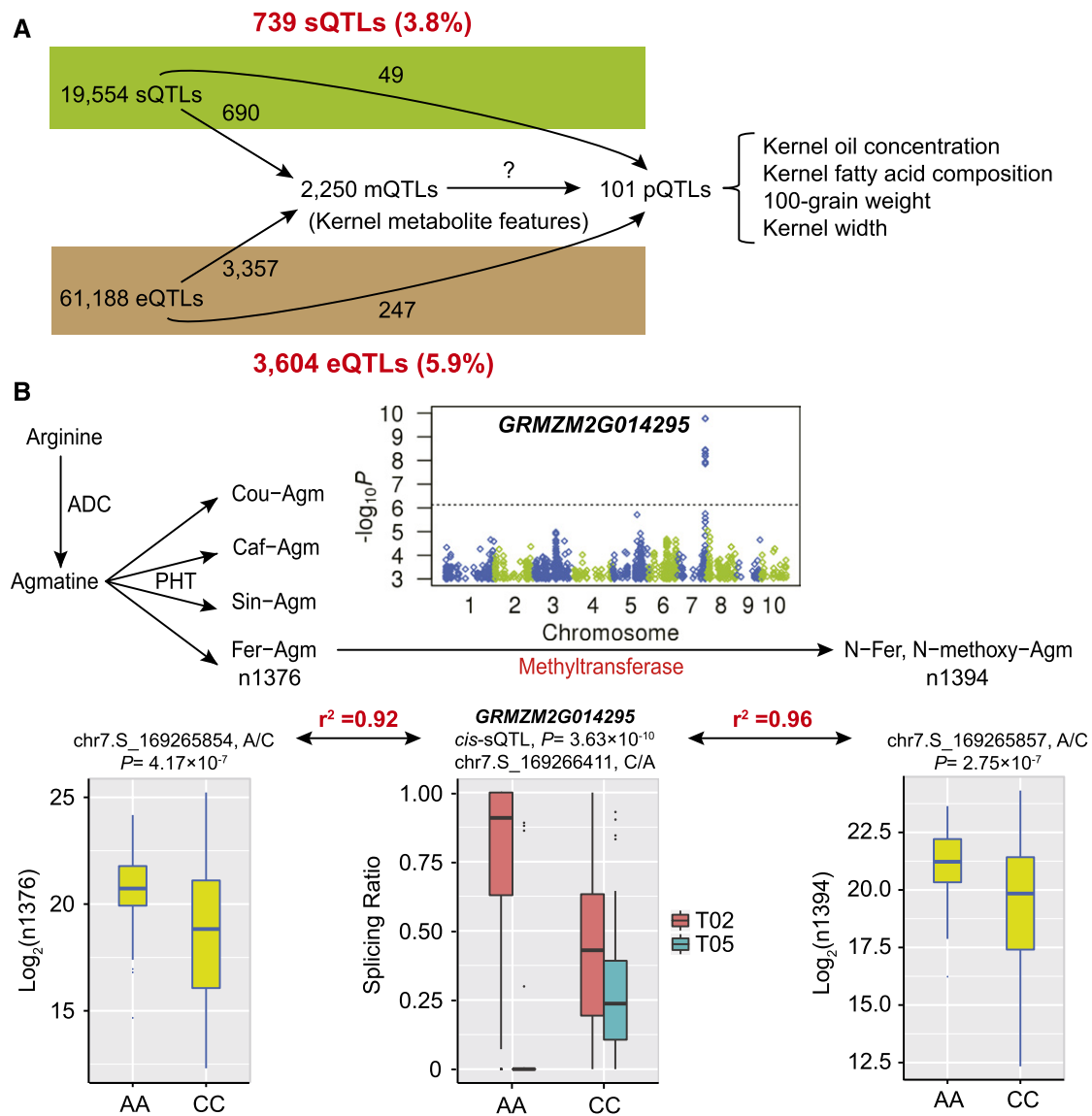
In one arginine metabolic pathway, arginine is first catalyzed by ADC (arginine decarboxylase) into agmatine, followed by the catalysis of PHT (putrescine hydroxycinnamoyltransferase) into various agmatine conjugates. The feruloyl-agmatine conjugates (n1376) are further modified by MT (methyltransferase) into n1394 (*N*-feruloyl, *N*-methoxyagmatine) (Wen et al., 2014). However, which gene encodes MT in the reaction is unclear. Significant mQTLs were detected for the n1376 ( $rs\#chr7.S\_169265854$ ,  $P = 4.17 \times 10^{-7}$ ) and n1394 ( $rs\#chr7.S\_169265857$ ,  $P = 2.75 \times 10^{-7}$ ) content in a previous metabolic study (Figure 11B; Wen et al., 2014). SNPs  $rs\#chr7.S\_169265854$  and  $rs\#chr7.S\_169265857$  were in strong LD ( $r^2 = 0.89$ ). Interestingly, we found that a significant *cis*-sQTL ( $rs\#chr7.S\_169266411$ ,  $P = 3.63 \times 10^{-10}$ ) for *GRMZM2G014295* was in strong LD with  $rs\#chr7.S\_169265854$  and  $rs\#chr7.S\_169265857$  ( $r^2 > 0.9$ ) (Figure 11B), potentially indicating that these three SNPs are in strong LD with a common causal variant. *GRMZM2G014295* encodes a methyltransferase. Given this, *GRMZM2G014295* is highly likely the underlying gene catalyzing the metabolic step from feruloyl-agmatine conjugates to *N*-feruloyl, *N*-methoxyagmatine. No significant difference in the expression level of *GRMZM2G014295* was detected at the *cis*-sQTL SNP. These results suggest that the metabolic content of n1394 and n1376 is most likely controlled via the splicing regulation of *GRMZM2G014295*.

### DISCUSSION

We assembled the transcriptome of immature kernels from 368 diverse maize inbred lines. A total of 48,855 novel transcripts from known genes were identified. The results showed that 50% (12,616 genes) of intron-containing genes undergo AS and display natural variation in the maize population. Considering that we only sampled one stage of kernel development, splicing variation could be widespread in natural populations. Some studies argued that much of the AS is not due to regulated splicing but represents splicing noise without biological consequences (Melamud and Moult, 2009; Pickrell et al., 2010a). Using a GWAS approach, we identified 19,554 unique sQTLs regulating the splicing variation in 6570 genes (representing 54% of mapped genes). The proportion of genes with sQTLs might be underestimated due to the use of a stringent Bonferroni-corrected threshold for claiming sQTL. Based on our results, we conclude that the widespread natural variation in AS is mostly under genetic control rather than resulting from splicing noise. Similar to the genetic architecture feature of gene expression level (Fu et al., 2013), splicing variation is under relatively simple genetic control. Moreover, we found that most sQTLs showed small isoform usage changes without involving major isoform switching between genotypes. Further studies of transcriptomes of other tissues at other developmental stages will provide a more complete picture of the genetic regulation of natural variation in AS.

Although increasing proteome diversity is one of the important postulated roles of AS, the extent of its influence on proteome diversity is far from being settled. Widespread natural variation in AS has been observed; however, the extent to which it leads to functional protein isoforms remains unclear. To this end, we compared the protein domain structure of isoforms differentially used at sQTLs. The sQTLs of the 1977 genes involved using isoform pairs exhibiting gained, lost, or exchanged domains relative to each other, potentially indicating that these sQTL-affected isoforms might have distinct protein functions. This functional divergence between isoforms might be underestimated because our current bioinformatic analysis of protein sequences of isoform pairs is highly dependent on the database of known protein domains. Many important protein domains or motifs might have not been characterized yet. Therefore, the functional divergence between isoforms encoded by the same gene might be widespread. Similar observations were reported in a recent comparative functional profiling of alternative isoforms in humans (Yang et al., 2016). Those authors found that AS can produce isoforms with vastly different interaction profiles, and alternative isoforms functionally behave like distinct proteins rather than minor variants of each other (Yang et al., 2016). Many genes encoding alternatively spliced isoforms with distinct functional characteristics have been identified in plants (Zhang and Mount, 2009; Posé et al., 2013; Wang et al., 2015). For example, the MADS box transcription factor gene *FLOWERING LOCUS M* (*FLM*) in *Arabidopsis* encodes two isoforms that regulate flowering in opposition, with *FLM-β* repressing flowering while *FLM-δ* activating flowering (Posé et al., 2013).

We provided compelling evidence that AS-coupled NMD, and miRNA-mediated mechanism and siPEP-mediated PEPi are frequently involved in sQTL regulation. Notably, the sQTLs



**Figure 11.** sQTLs Colocalize with Trait Associations.

**(A)** Overall summary of the colocalization of sQTLs with mQTLs and pQTLs.

**(B)** Pathway of degradation from arginine to agmatine. The derived conjugate Fer-Agm is further modified into N-Fer, N-methoxy-Agm. ADC, arginine decarboxylase; PHT, putrescine:hydroxycinnamoyltransferase. A cis-sQTL (rs#chr7.S\_169266411) detected at *GRMZM2G014295*, which encodes a methyltransferase, colocalized with mQTL for the content of Fer-Agm (rs#chr7.S\_169265854) and N-Fer, N-methoxy-Agm (rs#chr7.S\_169265857). Transcripts named “T” stand for known transcript.

of 1643 genes involved the differential use of isoform pairs that gain or lost NMD features. This result is consistent with recent results in Arabidopsis, in which NMD is a widespread mechanism for regulating gene expression, with 11 to 18% of alternatively spliced transcripts degraded by NMD (Kalyna et al., 2012; Drechsel et al., 2013). Previous studies have shown that AS typically produces transcripts that gained or lost miRNA target sites (Thatcher et al., 2014). We predicted miRNA targets against all sQTL-affected transcripts and found that the sQTLs of 16 genes involved differential use of alternative isoforms

carrying miRNA target sites. We should also note that due to the current limited miRNA annotation in maize only 70 high-confidence miRNAs were used in this analysis. A miRNA-mediated mechanism might be prevalently involved in sQTL regulation. In addition, siPEP-mediated PEPi is a recently proposed gene regulation mechanism (Seo et al., 2011a, 2013; Staudt and Wenkel, 2011). Our analysis in this study provided an overall evaluation of AS-coupled PEPi in maize. The results showed that the sQTLs of 161 transcription factors are potentially involved in regulation by PEPi. These results suggest

that AS functions as a central modulator in the gene regulatory network by coupling with different layers of gene regulation mechanisms, which allows both coordination and versatility in gene expression for plants to respond robustly and precisely to genetic and environmental perturbations.

The analysis in this study showed that only a small proportion of variants were simultaneously associated with total mRNA level and AS, indicating that the natural variation in overall mRNA level and AS tend to be independently regulated. Similar results were also observed in recent studies in humans, in which most sQTLs (74%) were not associated with gene expression levels (Li et al., 2016). Our further variant analysis for *cis*-sQTLs and *cis*-eQTLs revealed that different *cis*-regulatory sequences are preferentially used for the *cis*-regulation of AS and mRNA level, which could partially explain the relative genetic independence between overall mRNA level and AS. Notably, we identified both *cis*-sQTL and *cis*-eQTL at 676 genes, but the associated SNPs were not in LD, indicating that these genes contain independent *cis*-variants that independently regulate the mRNA levels and AS. These results suggested that plants may have evolved elaborate regulatory mechanisms to modulate mRNA levels and AS variation in response to developmental and environmental cues.

Previous studies have shown that mutations affecting splicing underlie a number of genes for plant growth, development, and responses to external cues (Cai et al., 1998; Isshiki et al., 1998; Seo et al., 2011b, 2012; Li et al., 2012). These individual gene studies illustrated that AS is an important factor determining phenotypic variation. However, an overall evaluation of the relative importance of AS in regulating phenotypic variation is generally lacking. The comparative analysis of sQTLs and previous marker-trait associations showed that 3.8% of mapped sQTLs were colocalized with previous association signals for metabolic and phenotypic traits of kernel scored in the same population. Notably, 83% of these colocalized sQTLs are not simultaneously associated with changes in overall mRNA levels. We applied the same colocalization analysis for the previous identified eQTLs with trait associations. The results showed that 5.9% of mapped eQTLs colocalized with trait associations (Figure 10A). This result suggested that AS, whose role has not been sufficiently appreciated in previous genetic studies, is as important as the gene expression levels in regulating plant phenotypes. Considering the relative genetic independence of AS from the overall gene expression level, our results demonstrate that AS represents an important source of trait associations and thus should be routinely considered in further genetic studies.

## METHODS

### Plant Materials

An association panel comprising 368 diverse maize inbred lines was used in this study (Yang et al., 2011; Li et al., 2013). Transcriptome sequencing was previously performed on the immature kernels at 15 DAP of these 368 inbred lines (Fu et al., 2013). The same transcriptome data were used for AS analysis in this study. The detailed information on the transcriptome sequencing can be found in the previous study (Fu et al., 2013).

Briefly, the association population was originally planted in two biological replications in the field (Fu et al., 2013). Five immature kernels from three to four ears in each block were collected at 15 DAP. The collected immature kernels in two replications were then bulked for total RNA extraction and sequencing.

### Transcriptome Assembly and Quantification

The paired-end (2 × 90 bp) libraries were filtered with FASTX-toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) to remove low-quality reads. Only reads with a quality score of more than 20 in at least 80% bases were retained and subsequently aligned to the B73 reference genome (version 3.31) using HISAT2 version 2.0.4 (Kim et al., 2015; Pertea et al., 2016), with a minimum intron size of 60 bp and maximum intron size of 50,000 bp. On average, 47.8 million reads were aligned to the genome and 41.9 million reads were aligned to one unique location in the genome. Only uniquely mapped reads were subsequently passed to StringTie version 1.3.0 (Pertea et al., 2015, 2016) for transcript assembly in each library using 63,230 annotated maize transcripts (AGPv3) as a reference transcriptome. The transcripts assembled from each library were combined into a unified set of transcripts using the merge function in StringTie. The discovery of novel junctions required at least 10 spanning reads and any new transcripts required to represent at least 10% of the total gene abundance in at least one library.

Known and novel transcripts were subsequently quantified in each inbred line using StringTie with default parameters. New transcripts from intergenic regions were not included in our analysis. According to a previous method (Thatcher et al., 2014), a simulation was performed to determine the expression threshold in fragments per kilobase of transcript per million mapped reads (FPKM) for transcript isoforms, and the details are shown in Supplemental Figure 6. Known and novel transcripts with expression less than 0.35 FPKM in <5% inbred lines were filtered out. To control the impact of transcript fragments, only transcripts with at least 70% long as the canonical transcript were kept for subsequent analysis. We subsequently calculated the splicing ratio of each isoform relative to the total transcript abundance of the gene, and the splicing ratio was used as the phenotype for genome-wide association analysis.

### sQTL Mapping

The genome v2 coordinates of 1.25 million SNPs from previous studies (Liu et al., 2017) were converted to genome v3 coordinates using Cross-Map version 2.0.5 (Zhao et al., 2014). Each of 1.25 million SNPs was tested for association with each splicing ratio. The population structure and kinship coefficients were estimated according to Fu et al. (2013). Principle component analysis was performed to detect the unknown confounding factors, and the first two PCs were included as covariates in association analysis. GWAS was performed to identify sQTLs using a mixed linear model accounting for population structure, family relatedness, and hidden confounders (Yu et al., 2006) with Tassel version 5.2.10 (Bradbury et al., 2007). A standard Bonferroni correction assumes the independence of all tests performed. Obviously, the SNPs used for GWAS are not independent because they are in LD at different levels. To determine how many independent statistical tests were actually performed in splicing GWAS, we performed LD analysis for the 1.25 million SNPs using PLINK (Purcell et al., 2007) with a window size of 50 SNPs and step size of 5 SNPs. We obtained 67,334 independent SNPs ( $r^2 < 0.1$ ). Therefore, a Bonferroni-corrected  $P < 7.43 \times 10^{-7}$  ( $0.05/n$ ,  $n = 67,334$ ) was used to define significant associations at  $\alpha = 0.05$ . We then performed two steps of filtering for the initial associations. First, for each gene detected with sQTL, we conducted LD analysis for the associated SNPs to identify independently associated SNPs. A unique sQTL was defined when the associated SNP was not in LD ( $r^2 < 0.1$ ) with any other associated SNPs on the same chromosome for the target gene. Second, we examined the

splicing ratio changes at each unique sQTL. Only sQTLs that exhibited more than 5% difference in splicing ratio were remained for subsequent analysis. We analyzed the relative positions of sQTLs and their associated genes and detected an obvious inflection point at the position of 20 kb (Figure 2B), which exhibited a similar distribution as that for the relative positions of eQTLs and their associated genes (Fu et al., 2013). Therefore, a sQTL was considered a *cis*-sQTL when the SNP was detected within 20 kb of the transcription start site or transcription end site of the target gene; otherwise, the sQTL was considered a *trans*-sQTL.

### sQTL Validation

To verify the sQTLs, we randomly selected 20 genes detected with local sQTLs and performed RT-PCR analysis to examine the relative differences in isoform expression between genotypes at the sQTL. To facilitate the RT-PCR analysis, the 20 genes were randomly selected from genes with two expressed isoforms. The materials used for RT-PCR assay were sampled from an independent experiment of the association panel that was grown in winter nursery (2015, Hainan, China) with two replications. For each inbred line, immature seeds collected from the two replications at 15 DAP were pooled together for total RNA extraction. For each genotypic class at the examined sQTL, five inbred lines carrying the same SNP allele were randomly selected to be bulked as a biological replicate, and three biological replicates were used in the RT-PCR analysis. Two micrograms of total RNA was reverse transcribed using a random primer and reverse transcriptase (Promega) following the manufacturer's instructions. cDNA was then analyzed by amplification of *Tubulin* (control) and transcript isoforms for each gene. Reactions were analyzed by agarose gel electrophoresis and ethidium bromide staining. The primers used in the RT-PCR assay can be found in Supplemental Data Set 12.

### Gene Function Enrichment Analysis

To test whether genes detected with sQTLs share common functional features, GO term and pathway enrichment analyses were performed using a hypergeometric test. GO terms were determined by the web toolkit agriGO (Du et al., 2010). GO categories that contain at least five genes were considered significantly enriched with an false discovery rate-corrected  $P < 0.05$ . Similarly, to test whether genes detected with sQTLs were enriched in specific pathways, MapMan database (Thimm et al., 2004) was used to categorize the pathways. Pathways containing at least five genes were considered significantly overrepresented with BH-corrected  $P < 0.05$ . In addition, a set of splicing-related genes from the splicing-related gene database in plantGDB (Duvick et al., 2008) and spliceosome genes from the KEGG database (Kanehisa et al., 2012) were also used for functional annotation.

### AS Event Characterization

To characterize the type of splicing events affected by sQTLs, pairwise comparison of sQTL-associated transcripts within each regulated gene was performed using the ASTALAVISTA tool (Foissac and Sammeth, 2007). ASTALAVISTA extracts and classifies the different splicing events based on the transcript model and splice junctions. The AS events were classified into (1) IR, (2) AA, (3) AD, (4) ES, and (5) other events. The "other events" represent complex AS events comprising duplicated IR, ES, AA, and AD events or a combination of those.

### Conserved Domain Prediction and PEPi Analysis

PfamScan was used to search protein domains against the Pfam HMM library based on HMMER3 (Eddy, 2011) for all novel and known tran-

scripts, using protein sequences as input. The Pfam database (Finn et al., 2016) (version 30.0) contains 16,306 protein families. ORFs for all novel and known transcripts were searched from authentic start codon for each gene. For transcripts that did not cover the authentic start codon, the longest ORF was used to search protein domains. After filtering at  $E$ -value  $< 1 \times 10^{-3}$ , resulting domain hits for novel and known transcripts were then compared in a pairwise manner. The annotation information from PlantTFDB (Jin et al., 2014), GrassTFDB (Yilmaz et al., 2009), and MapMan (Thimm et al., 2004) database was integrated as database of transcription factors. For each transcription factor regulated by sQTLs, the domain configuration of transcript isoforms was compared in a pairwise manner to determine whether the isoform pair involved a PEPi mechanism. Read coverage was obtained by *genomecov* function in bedtools (Quinlan and Hall, 2010) with *-dz -split* parameters.

### NMD Analysis

The coding sequences (CDSs) of novel transcripts were first predicted using EMBOSS version 6.4.0 (Rice et al., 2000), and the ORF translating from authentic start codon was considered the coding region. To determine whether the isoform regulated by sQTL was a putative NMD target, the transcript structure of isoform pairs was compared. Transcript isoforms without IR events carrying a PTC located at least 50 nucleotides upstream of an exon splice junction, 3'-UTRs longer than 350 bp, introns within 3'-UTRs, and upstream open reading frames greater than 35 amino acids within 5'-UTRs (Kalyna et al., 2012; Drechsel et al., 2013) were marked as potential NMD candidates.

### miRNA Target Analysis

To determine whether sQTL-affected isoforms involve miRNA regulation mechanism, the potential miRNA target sites of isoforms were predicted using targetfinder.pl (Fahlgren and Carrington, 2010) with miRBase Release 22 database, which contains 325 maize miRNAs. To reduce false positives of miRNA target prediction, only 70 miRNAs that were annotated as high-confidence mature miRNAs were used for prediction. Using a default miRNA/target complementarity score of 4, miRNA target sites were obtained for each transcript.

### Variant Enrichment Analysis

The effect of each variant was annotated by the SnpEff program (Cingolani et al., 2012) to derive functional consequences based on different genomic regions and effect consequences. The hypergeometric test (*phyper* function in R) was performed to test the enrichment of the ratio of different types of SNP effects compared with background.

### Modified 5'-RACE

To examine the cleavage sites in the *ZmGRF8* transcripts, we performed modified 5'-RNA ligation-mediated (RLM) RACE following the manufacturer's instructions of the FirstChoice RLM-RACE kit (Invitrogen). The cDNA templates from developing kernels of B73 at 15 DAP were amplified through two rounds of PCR with the universal sense or antisense primers provided in the kit and two gene-specific primers. The PCR products were cloned into the pEASY vector (TransGen) and sequenced.

### Knockout of *ZmGRP1*

A pCAMBIA-derived CRISPR/Cas9 binary vector with two gRNA expression cassettes targeting two adjacent sites of *ZmGRP1* was generated according to previously described protocol (Xing et al., 2014). The construct was transformed into *Agrobacterium tumefaciens* strain EHA105,



and then *Agrobacterium*-mediated method was used to transform the immature embryos of receptor line. A total of six independent T0 transgenic plants were obtained. In order to identify the positive CRISPR-Cas9 knockout lines, the PCR amplicons encompassing the gRNA-targeted sites for each of the transgenic plants were separated on agarose gels and cloned in the pMD-18T vector (Takara). At least eight clones per PCR product were sequenced. After PCR amplification and sequencing, two T0 plants with homozygous deletions at the target sites were identified (referred to as KO#2 and KO#5). The homozygous T0 plant KO#2 was self-pollinated to get T1 progenies. The resulting T1 progenies, together with the wild type, were planted and further genotyped by PCR and sequencing to confirm the presence of mutations at the target sites. Immature kernels of T1 and wild-type plants were collected at 15 DAP for mRNA sequencing. The primers used are listed in Supplemental Data Set 12.

### mRNA Sequencing and Exon Usage Differential Analysis

Total RNA was isolated and purified using RNAprep pure plant kits (Tiangen Biotech). Libraries of mRNA were prepared and sequenced using HiSeq 150-bp paired-end Illumina RNA-Seq protocol, with three and four biological replicates for KO#2 and the wild type, respectively. After filtering out reads with low sequencing quality, ~44.5 million reads were maintained in each sample (Supplemental Data Set 13). The resulting sequences were aligned to the B73 genome sequence v3 (AGPv3.31) with HISAT2 (Kim et al., 2015; Pertea et al., 2016) using a minimum intron size of 60 bp and a maximum intron size of 50,000 bp. Only uniquely mapped reads were kept for subsequent analysis. On average, 37 million reads were uniquely mapped onto the reference genome (Supplemental Data Set 13). Genome-matched reads were assembled for each sample with Cufflinks (Trapnell et al., 2012), and the merged transcripts from all samples were used to count reads mapped onto each exon. Differential exon usage between KO#2 and the wild type were tested using DEXSeq packages, which uses generalized linear models and offers reliable control of false discoveries by taking biological variation into account (Anders et al., 2012). The P value for each exon was obtained by a  $\chi^2$  likelihood ratio test and adjusted using Benjamini-Hochberg method (Anders et al., 2012).

### RT-qPCR Analysis

For sQTL validation, four inbred lines for each genotype at sQTL were selected for RT-qPCR validation. For differentially used exon validation, the same samples used for RNA-seq were used for RT-qPCR validation. Total RNA (2  $\mu$ g) was reverse transcribed using a random primer and reverse transcriptase (Promega) following the manufacturer's instructions. RT-qPCR was performed on an ABI 7500 real-time PCR system using the SYBR Premix Ex Taq II kit (Takara), and measurements were obtained using the comparative CT ( $2^{-\Delta\Delta CT}$ ) method relative quantification method (Schmittgen and Livak, 2008). *Tubulin* was used as an internal control. All data are based on three independent biological replicates, each with three technical replicates. The primers used are listed in Supplemental Data Set 12.

### Accession Numbers

The raw sequence of RNA-seq for *ZmGRP1* was deposited into Sequence Read Archive of National Center for Biotechnology Information under accession number SRP132125. Sequence data from this article can be found in the GenBank/EMBL databases under the following accession numbers: *ZmGRF8*, *GRMZM2G041223*; *su2*, *GRMZM2G348551*; *ZmMADS1*, *GRMZM2G171365*; *Cys2*, *GRMZM2G005887*; *ZmGRP1*, *GRMZM2G080603*; *AtGRP7*, *AT2G21660*; and *AtGRP8*, *AT4G39260*.

### Supplemental Data

**Supplemental Figure 1.** RT-PCR Validation of sQTLs.

**Supplemental Figure 2.** Distribution of sQTL-Associated Alternative Splicing Events.

**Supplemental Figure 3.** Permutation Tests to Examine Whether the Observed Results Were Expected by Chance or Not (1000 Permutations).

**Supplemental Figure 4.** Sequence Analysis of CRISPR-Cas9 Knockout Line KO#2 of *ZmGRP1*.

**Supplemental Figure 5.** RT-qPCR Results of Differentially Used Exons for Knockout of *ZmGRP1*.

**Supplemental Figure 6.** Determination of Expression Cutoff.

**Supplemental Data Set 1.** sQTL Mapping Summary.

**Supplemental Data Set 2.** Pathway Enrichment Results for sQTL-Regulated Genes.

**Supplemental Data Set 3.** Gene Ontology Enrichment Results for sQTL-Regulated Genes.

**Supplemental Data Set 4.** Protein Domain Comparison for sQTL-Associated Isoforms.

**Supplemental Data Set 5.** NMD Analysis Summary.

**Supplemental Data Set 6.** miRNA Target Site Prediction Results.

**Supplemental Data Set 7.** Summary for Genes with Independent Genetic Control of *cis*-sQTL and *cis*-eQTL.

**Supplemental Data Set 8.** Summary for *Trans*-Acting Splicing Regulators.

**Supplemental Data Set 9.** Differential Exon Usage Results for Knockout of *ZmGRP1*.

**Supplemental Data Set 10.** GO Enrichment for Genes with Significant Exon Usage Differences by Knockout of *ZmGRP1*.

**Supplemental Data Set 11.** Summary for sQTL Colocalizing with Trait Associations.

**Supplemental Data Set 12.** Summary of Primer Sequences Used in the Study.

**Supplemental Data Set 13.** mRNA-Seq Mapping Statistics for Knockout of *ZmGRP1*.

### ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (31421005), the National Key Research and Development Program of China (2016YFD100404), the Recruitment Program of Global Experts, and the Fundamental Research Funds for the Central Universities.

### AUTHOR CONTRIBUTIONS

Q.C., Y.H., and H.L. contributed equally to this work. F.T. and X.Y. designed the project. Q.C., H.L., J.S., and X.W. performed data analyses. Y.H., Q.C., B.Z., W.L., J.T., and Y.L. performed experiments. F.T., Q.C., and X.Y. wrote the manuscript. F.T., X.Y., and J.Y. supervised the work.

Received February 6, 2018; revised May 23, 2018; accepted June 27, 2018; published July 2, 2018.

### REFERENCES

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**: 2008–2017.

- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V.** (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**: 76–84.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S.** (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.
- Brogna, S., McLeod, T., and Petric, M.** (2016). The meaning of NMD: translate or perish. *Trends Genet.* **32**: 395–407.
- Cai, X.L., Wang, Z.Y., Xing, Y.Y., Zhang, J.L., and Hong, M.M.** (1998). Aberrant splicing of intron 1 leads to the heterogeneous 5' UTR and decreased expression of *waxy* gene in rice cultivars of intermediate amylose content. *Plant J.* **14**: 459–465.
- Chang, Y.-F., Imam, J.S., and Wilkinson, M.F.** (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**: 51–74.
- Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M.** (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80–92.
- Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., Zhang, X., Jin, M., Zhao, M., and Yan, J.** (2017). The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol. J.* **15**: 1250–1263.
- Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rättsch, G., and Wachter, A.** (2013). Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. *Plant Cell* **25**: 3726–3742.
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z.** (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**: W64–W70.
- Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M.D., Lawrence, C.J., Lushbough, C., and Brendel, V.** (2008). PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* **36**: D959–D965.
- Eddy, S.R.** (2011). Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**: e1002195.
- Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J., and Schaal, H.** (2013). Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* **19**: 96–102.
- Fahlgren, N., and Carrington, J.C.** (2010). miRNA target prediction in plants. *Methods Mol. Biol.* **592**: 51–57.
- Finn, R.D., et al.** (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**: D279–D285.
- Foissac, S., and Sammeth, M.** (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* **35**: W297–W299.
- Fu, J., et al.** (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat. Commun.* **4**: 2832.
- Gan, X., et al.** (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**: 419–423.
- Geuens, T., Bouhy, D., and Timmerman, V.** (2016). The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135**: 851–867.
- Göhring, J., Jacak, J., and Barta, A.** (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in *Arabidopsis*. *Plant Cell* **26**: 754–764.
- GTE Consortium** (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660.
- Huang, J., Gao, Y., Jia, H., Liu, L., Zhang, D., and Zhang, Z.** (2015). Comparative transcriptomics uncovers alternative splicing changes and signatures of selection from maize improvement. *BMC Genomics* **16**: 363.
- Isshiki, M., Morino, K., Nakajima, M., Okagaki, R.J., Wessler, S.R., Izawa, T., and Shimamoto, K.** (1998). A naturally occurring functional allele of the rice *waxy* locus has a GT to TT mutation at the 5' splice site of the first intron. *Plant J.* **15**: 133–138.
- Iwakawa, H.O., and Tomari, Y.** (2015). The functions of microRNAs: mRNA decay and translational repression. *Trends Cell Biol.* **25**: 651–665.
- Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J.** (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **42**: D1182–D1187.
- Kalyna, M., et al.** (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* **40**: 2454–2469.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.** (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**: D109–D114.
- Kim, D., Langmead, B., and Salzberg, S.L.** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**: 357–360.
- Kim, S.H., Koroleva, O.A., Lewandowska, D., Pendle, A.F., Clark, G.P., Simpson, C.G., Shaw, P.J., and Brown, J.W.** (2009). Aberrant mRNA transcripts and the nonsense-mediated decay proteins UPF2 and UPF3 are enriched in the *Arabidopsis* nucleolus. *Plant Cell* **21**: 2045–2057.
- Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J.** (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**: 153–165.
- Kurihara, Y., et al.** (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **106**: 2453–2458.
- Li, H., et al.** (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **45**: 43–50.
- Li, Q., Lin, Y.C., Sun, Y.H., Song, J., Chen, H., Zhang, X.H., Sederoff, R.R., and Chiang, V.L.** (2012). Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA* **109**: 14699–14704.
- Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K.** (2016). RNA splicing is a primary link between genetic variation and disease. *Science* **352**: 600–604.
- Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., and Yan, J.** (2017). Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol. Plant* **10**: 414–426.
- Mandadi, K.K., and Scholthof, K.-B.G.** (2015). Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *Plant Cell* **27**: 71–85.
- Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M.** (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* **22**: 1184–1195.
- Mei, W., Liu, S., Schnable, J.C., Yeh, C.T., Springer, N.M., Schnable, P.S., and Barbazuk, W.B.** (2017). A comprehensive analysis of alternative splicing in paleopolyploid maize. *Front. Plant Sci.* **8**: 694.

- Melamud, E., and Moul, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37**: 4873–4886.
- Monlong, J., Calvo, M., Ferreira, P.G., and Guigó, R. (2014). Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.* **5**: 4698.
- Naftelberg, S., Schor, I.E., Ast, G., and Kornblihtt, A.R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**: 165–198.
- Pertea, M., Mount, S.M., and Salzberg, S.L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinformatics* **8**: 159.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**: 290–295.
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**: 1650–1667.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010b). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**: 768–772.
- Pickrell, J.K., Pai, A.A., Gilad, Y., and Pritchard, J.K. (2010a). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **6**: e1001236.
- Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* **503**: 414–417.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**: 559–575.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Reddy, A.S.N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* **25**: 3657–3683.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**: 276–277.
- Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**: 1101–1108.
- Seo, P.J., Hong, S.Y., Kim, S.G., and Park, C.M. (2011a). Competitive inhibition of transcription factors by small interfering peptides. *Trends Plant Sci.* **16**: 541–549.
- Seo, P.J., Kim, M.J., Ryu, J.Y., Jeong, E.Y., and Park, C.M. (2011b). Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* **2**: 303.
- Seo, P.J., Park, M.-J., Lim, M.-H., Kim, S.-G., Lee, M., Baldwin, I.T., and Park, C.-M. (2012). A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. *Plant Cell* **24**: 2427–2442.
- Seo, P.J., Park, M.J., and Park, C.M. (2013). Alternative splicing of transcription factors in plant responses to low temperature stress: mechanisms and functions. *Planta* **237**: 1415–1424.
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., Ma, Y., Liu, T., Kong, L.A., Peng, D.L., and Tian, Z. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* **26**: 996–1008.
- Staiger, D., and Brown, J.W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* **25**: 3640–3656.
- Staudt, A.C., and Wenkel, S. (2011). Regulation of protein function by ‘microProteins’. *EMBO Rep.* **12**: 35–42.
- Streitner, C., Hennig, L., Korneli, C., and Staiger, D. (2010). Global transcript profiling of transgenic plants constitutively overexpressing the RNA-binding protein AtGRP7. *BMC Plant Biol.* **10**: 221.
- Streitner, C., Köster, T., Simpson, C.G., Shaw, P., Danisman, S., Brown, J.W., and Staiger, D. (2012). An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* **40**: 11240–11255.
- Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W. (2012). Alternative splicing in plants—coming of age. *Trends Plant Sci.* **17**: 616–623.
- Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* **8**: 14519.
- Thatcher, S.R., Zhou, W., Leonard, A., Wang, B.B., Beatty, M., Zastrow-Hayes, G., Zhao, X., Baumgarten, A., and Li, B. (2014). Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell* **26**: 3472–3487.
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**: 914–939.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**: 562–578.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Wang, Z., Ji, H., Yuan, B., Wang, S., Su, C., Yao, B., Zhao, H., and Li, X. (2015). ABA signalling is fine-tuned by antagonistic HAB1 variants. *Nat. Commun.* **6**: 8138.
- Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**: 3438.
- Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., and Chen, Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**: 327.
- Yang, X., et al. (2016). Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**: 805–817.
- Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J., and Yan, J. (2014). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* **10**: e1004573.
- Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J., and Yan, J. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breeding* **28**: 511–526.

- Yilmaz, A., Nishiyama, M.Y., Jr., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J., and Grotewold, E.** (2009). GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol.* **149**: 171–180.
- Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S.** (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**: 203–208.
- Zhang, G., et al.** (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **20**: 646–654.
- Zhang, X.N., and Mount, S.M.** (2009). Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. *Plant Physiol.* **150**: 1450–1458.
- Zhang, X., Joehanes, R., Chen, B.H., Huan, T., Ying, S., Munson, P.J., Johnson, A.D., Levy, D., and O'Donnell, C.J.** (2015). Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* **47**: 345–352.
- Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.P., and Wang, L.** (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**: 1006–1007.