



# CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Khokhar, W., Hassan, M., Reddy, A., Chaudhary, S., Jabre, I., Byrne, L. and Syed, N. (2019) Genome-wide identification of splicing quantitative trait loci (sQTLs) in diverse ecotypes of *Arabidopsis thaliana*. *Frontiers in Plant Science*, 10 (1160). ISSN 1664-462X.

Link to official URL (if available):

<https://dx.doi.org/10.3389/fpls.2019.01160>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: [create.library@canterbury.ac.uk](mailto:create.library@canterbury.ac.uk)





# Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse Ecotypes of *Arabidopsis thaliana*

Waqas Khokhar<sup>1</sup>, Musa A. Hassan<sup>2,3</sup>, Anireddy S. N. Reddy<sup>4</sup>, Saurabh Chaudhary<sup>1</sup>, Ibtissam Jabre<sup>1</sup>, Lee J. Byrne<sup>1</sup> and Naeem H. Syed<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Kranthi Kiran Mandadi,  
Texas A&M University,  
United States

### Reviewed by:

William Brad Barbazuk,  
University of Florida,  
United States  
Stephen M. Mount,  
University of Maryland,  
College Park, United States  
Renesh Bedre,  
Texas A&M University,  
United States

### \*Correspondence:

Naeem Hasan Syed  
naeem.syed@canterbury.ac.uk

### Specialty section:

This article was submitted to  
Plant Physiology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 15 March 2019

**Accepted:** 26 August 2019

**Published:** 03 October 2019

### Citation:

Khokhar W, Hassan MA, Reddy ASN, Chaudhary S, Jabre I, Byrne LJ and Syed NH (2019) Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse Ecotypes of *Arabidopsis thaliana*. *Front. Plant Sci.* 10:1160. doi: 10.3389/fpls.2019.01160

<sup>1</sup> School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, United Kingdom, <sup>2</sup> Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, <sup>3</sup> Centre for Tropical Livestock Genetics and Health, University of Edinburgh, Edinburgh, United Kingdom, <sup>4</sup> Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO, United States

Alternative splicing (AS) of pre-mRNAs contributes to transcriptome diversity and enables plants to generate different protein isoforms from a single gene and/or fine-tune gene expression during different development stages and environmental changes. Although AS is pervasive, the genetic basis for differential isoform usage in plants is still emerging. In this study, we performed genome-wide analysis in 666 geographically distributed diverse ecotypes of *Arabidopsis thaliana* to identify genomic regions [splicing quantitative trait loci (sQTLs)] that may regulate differential AS. These ecotypes belong to different microclimatic conditions and are part of the relict and non-relict populations. Although sQTLs were spread across the genome, we observed enrichment for *trans*-sQTL (*trans*-sQTLs hotspots) on chromosome one. Furthermore, we identified several sQTL (911) that co-localized with trait-linked single nucleotide polymorphisms (SNP) identified in the *Arabidopsis* genome-wide association studies (AraGWAS). Many sQTLs were enriched among circadian clock, flowering, and stress-responsive genes, suggesting a role for differential isoform usage in regulating these important processes in diverse ecotypes of *Arabidopsis*. In conclusion, the current study provides a deep insight into SNPs affecting isoform ratios/genes and facilitates a better mechanistic understanding of trait-associated SNPs in GWAS studies. To the best of our knowledge, this is the first report of sQTL analysis in a large set of *Arabidopsis* ecotypes and can be used as a reference to perform sQTL analysis in the Brassicaceae family. Since whole genome and transcriptome datasets are available for these diverse ecotypes, it could serve as a powerful resource for the biological interpretation of trait-associated loci, splice isoform ratios, and their phenotypic consequences to help produce more resilient and high yield crop varieties.

**Keywords:** splicing quantitative trait loci (sQTL), *Arabidopsis thaliana*, alternative splicing, isoform usage, GWAS, adaptation

## INTRODUCTION

Plants have evolved diverse and efficient genetic and physiological strategies to cope with environmental fluctuations. For an appropriate response, plants employ different regulatory mechanisms that can modulate genomic architecture and transcriptome composition to generate phenotypic diversity, allowing them to engender appropriate responses and occupy diverse niches (Frachon et al., 2018). The majority of protein-coding genes (~90%) in plants contain introns, which must be removed by a process called pre-mRNA splicing to produce mature messenger RNAs (mRNAs) for translation. Due to differential exon and splice site usage, ~70% of plant genes can be alternatively spliced (AS) to generate few to thousands of structurally and functionally different mRNA isoforms with different fates (Filichkin et al., 2010; Marquez et al., 2012; Shen et al., 2014; Klepikova et al., 2016; Zhang et al., 2017). Interestingly, the majority of splicing regulator genes in plants are subject to extensive AS and change the profile of their splicing patterns in response to various environmental stresses (Palusa et al., 2007; Zhang et al., 2013; Calixto et al., 2018). AS events in some plant-specific SR genes are highly conserved from single-cell green alga *Chlamydomonas reinhardtii* or moss *Physcomitrella patens* to Arabidopsis, suggesting the importance of their regulation in plants (Iida and Go, 2006; Kalyna et al., 2006). Plants employ AS to fine-tune their physiology and metabolism to maintain a balance between carbon fixation and resource allocation under normal as well as biotic and/or abiotic stress conditions such as pathogen infection, temperature, salt, drought, wounding, and light (Feng et al., 2015; Huang et al., 2017; Calixto et al., 2018; Filichkin et al., 2018; Hartmann et al., 2018; Liu et al., 2018; Seaton et al., 2018). For example, short read (Illumina) and long read (Iso-seq) RNA-Seq data from poplar leaf, root, and stem xylem tissues under drought, salt, and temperature fluctuations revealed changes in AS profiles that modulate plant transcriptome under abiotic stresses (Filichkin et al., 2018). Moreover, intron retention (IR) was found to be the predominant and variable type of AS across all treatments and tissue types (Filichkin et al., 2018). In wheat, global profiling of the AS landscape responsive to drought, heat, and their combination showed significant AS variation (Liu et al., 2018). Recent studies have shown that salt stress, high temperature, disease, and cold-stress alter AS patterns in Arabidopsis, grape, rice, and soybean (Liu et al., 2013; Feng et al., 2015; Filichkin et al., 2015; Lee et al., 2016; Jiang et al., 2017; Calixto et al., 2018). Similarly, AS plays a key role in biotic stress responses; for example, data from soybean show that PsAvr3c, a *Phytophthora sojae* pathogen effector, can manipulate host spliceosomal machinery to shift splicing profiles and overcome the host immune system (Huang et al., 2017). However, it is not clear to what extent sequence variation influences splicing variation and whether chromatin environment also contributes towards it, in a condition- and stress-dependent manner (Syed et al., 2012; Pajoro et al., 2017; Jabre et al., 2019).

Genome and transcriptome sequencing in multiple accessions of Arabidopsis revealed that genetic variations influence the expression and splicing of several genes, including stress-responsive genes (Gan et al., 2011). For instance, a strong

association between genetic variations and spliced isoform accumulation was observed in sunflower (Smith et al., 2018). Interestingly, a significant proportion of splicing variation was associated with variants that harbor *trans*-QTLs, of which the majority were associated with spliceosomal proteins (Smith et al., 2018). Further examination of splicing variation in the wild and cultivated sunflowers revealed that higher frequency of AS was triggered during the domestication process. Some of the genetic variations can also influence important life-history traits like flowering and may have an influence on the geographical distribution of different accessions. For example, insertion polymorphisms in the first intron of the *FLOWERING LOCUS M (FLM)* gene influence AS and accelerate flowering in a temperature-dependent manner in many accessions of Arabidopsis (Lutz et al., 2015). Taken together, structural variation in the *FLM* gene can change ratios of different splice variants and influence a highly adaptive trait-like flowering in Arabidopsis (Lutz et al., 2015).

Insight into trait-associated genetic variants and their distribution pattern can delineate the mechanisms of genome regulation (Alonso-Blanco et al., 2016). Since AS can increase transcriptome/proteome complexity, the genetic underpinnings of natural sequence variations and AS are strongly associated with each other. Genetic variants, such as single nucleotide polymorphisms (SNPs), can substantially regulate the expression of transcript isoforms by modulating splice sites, which can impact phenotypic diversity and susceptibility to diseases in humans (Singh and Cooper, 2012; Takata et al., 2017). Recent studies showed that 22% of SNPs that are associated with different human diseases affect splicing (Qu et al., 2017; Park et al., 2018). The advances in RNA-Seq and genotyping have augmented the opportunities to monitor genetic variants and quantify transcriptomic features that allow us to understand the genetic landscapes of AS (Smith et al., 2018). Recent studies in animals and plants have elucidated the association of genetic variants with trait-associated loci at a population level (Monlong et al., 2014; Chen et al., 2018). In Arabidopsis, many studies have identified expression quantitative trait loci (eQTL) to explain trait-associated loci (Zhang et al., 2011). However, there have been very few studies on the genome-wide investigation of genetic variants affecting splicing patterns termed as splicing quantitative trait loci (sQTLs) in Arabidopsis (Yoo et al., 2016). To illuminate the role of genetic variations on AS in a large collection of highly diverse Arabidopsis lines, we sought to map sQTLs influencing AS. sQTLs spread across the genome can either act in *cis*- to disrupt the splicing of a proximal pre-mRNA by modulating splicing factors binding affinity to the pre-mRNA or in *trans* by regulating the splicing of distal pre-mRNA through altered expression of splicing regulators (Yoo et al., 2016; Qu et al., 2017).

We have used 666 diverse natural inbred Arabidopsis accessions including 'relicts' that occupied postglacial Eurasia first and were later invaded by 'non-relicts', which demographically spread along the east-west axis of Eurasia owing to its higher latitudinal regional diversity, human disturbance and climatic pressure (François et al., 2008; Alonso-Blanco et al., 2016). These accessions are of immense significance as they hold a huge amount of diversity and their

expansion leaves traces of admixture in the north and south of the species range that facilitated colonization to new habitats (Lee et al., 2017). In order to illuminate the relationship between splicing variants, phenotypic diversity and geographical distribution of these lines, we have performed sQTL analysis to reveal the functional impact of genetic variations on AS and adaptive consequences (Han et al., 2017). This analysis will provide a solid platform in the form of a useful and well-enriched dataset for sQTL in Arabidopsis to develop more resilient plant species in the face of climatic challenges to crop production.

## MATERIALS AND METHODS

### Genotype Datasets and Quality Control

High-quality genetic variant (SNPs) data for 1,135 globally distributed natural inbred lines of Arabidopsis representing relicts (accessions that hold ancestral habitats) and non-relicts (accessions range from native Eurasia to recently colonized North America) were downloaded in variant calling format (vcf) from the 1001 Genomes Data Centre (<https://1001genomes.org/data/GMI-MPI/releases/v3.1/>) (Alonso-Blanco et al., 2016). The genetic variants were pre-processed using stringent filtering criteria as follows: (i) having at least two genotypes across accessions, and (ii) each genotype has at least five occurrences across all accessions (Yang et al., 2017). Genotypes that occurred in less than five samples were converted to NA values to avoid their consideration in linkage analysis. The pre-processed high-quality genetic variant data was then used as input in mapping sQTLs (Supplementary Figure 1).

### RNA-Sequencing Analysis

Single-ended 100 bp long RNA sequencing (RNA-Seq) reads, generated by Illumina HiSeq 2500 platform, for 727 ecotypes of Arabidopsis (without biological replicates) were downloaded from GEO dataset under accession number GSE80744 and SRA study SRP074107 (Kawakatsu et al., 2016). Read quality assessment was performed using FastQC and reads with Phred score < 20 were removed *via* Trim Galore version 0.5.0 (Andrews, 2010; Krueger, 2015). Transcript abundance level in terms of transcripts per million (TPM), was estimated for 82,910 isoforms in 34,212 genes using Arabidopsis reference transcript dataset (AtrTD2) and Salmon (Zhang et al., 2015; Patro et al., 2017) (Supplementary Dataset 1). The filtered RNA-seq reads were also mapped to the TAIR10 genome assembly using the STAR aligner version 2.7.0e with modified parameters (Supplementary Dataset 1) (Dobin et al., 2013). Transcripts were assembled and assemblies were merged using StringTie (Pertea et al., 2015) (Supplementary Dataset 4). The merged assembly was then used as a reference gene model to perform the second assembly to generate expression dataset for 124,422 expressed transcripts using StringTie (Pertea et al., 2015). The TPM values for both (transcriptome- and genome-based) expression datasets were then used to compute splicing ratios of each isoform for all genes. Genes with less than two isoforms or splicing variability <0.01 were filtered out using the core functionality of ulfasQTL method (Yang et al., 2017).

### Identification of sQTLs

The genotype data from 1,135 and RNA-seq data from 727 Arabidopsis accessions were initially processed to filter 666 accessions that have both genotype and transcriptomic data (Supplementary Table 1; Supplementary Figure 1). The processed genetic variants and expression datasets were used to perform a genome-wide scan for sQTLs using the ulfasQTL package (v 0.1) – a composite sQTL analysis package that takes expression and genotype dataset to test splicing QTLs at genome-wide scale (Yang et al., 2017). It uses the core functionality of sQTLseeker approach, which is a multivariate model and calculates splicing ratios variability of a gene across samples using a distance-based approach. It estimates intra- and inter-genotype splicing variability using a non-parametric analog to the ANOVA (Monlong et al., 2014). ulfasQTL identifies and outputs a list of significant sQTLs ( $p$ -value  $\leq 0.05$ ) and their cognate genes across the genome. We used sQTL cognate genes to derive modes of AS events present in these genes using SUPPA version 2.2.1 (Alamancos et al., 2014; Alamancos et al., 2015), which provides an estimate of the inclusion level of AS events across all samples.

### Colocalization of sQTL With GWAS Hits

The list of trait-associated loci was downloaded from the Arabidopsis genome-wide association studies (AraGWAS) catalog, which is a manually curated and standardized database that holds GWAS results for 167 publicly available phenotypes of Arabidopsis (Togninalli et al., 2018). It contains around 222,000 SNP-trait associations (GWAS hits), of which 3,887 are highly significant ( $p$ -value <  $10^{-4}$ ). The list of unique sQTLs was then matched with trait-associated loci (identical SNPs) to identify the significant association with important phenotypic traits associated sQTLs.

### Gene Enrichment Analysis

Functional enrichment analysis was performed on the parent genes with significant sQTL using Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.8) with default parameters that work on the principle of Fisher exact test for statistical analysis (Huang et al., 2009). The gene ontology (GO) terms [biological process (BP), molecular function (MF), and cellular components (CC)] were identified to provide biological insights into the significant sQTLs using false discovery rate (FDR)  $\leq 0.05$ .

### Functional Annotation of sQTL and Non-sQTL SNPs

Publicly available functionally annotated SNPs dataset for 666 Arabidopsis accessions was downloaded and overlapped with sQTL-SNPs to obtain functional annotation of sQTLs and termed the other unmatched SNPs as non-sQTL SNPs. The SNPs were further classified into exonic-SNPs (nonsense, start-loss, frameshift, splice site, missense, synonymous, splice region, 5-UTR, 3-UTR, and non-coding exon variants) and non-exonic SNPs (intron and intergenic variants).

## RESULTS

### Majority of Splicing Events in *A. thaliana* Are Regulated as *Trans*

We performed genome-wide sQTL analysis using transcriptomic and genomic datasets and identified 6,406 and 6484 unique sQTLs that are associated with 6,129 and 7653 non-redundant genes, respectively (Table 1; Supplementary Tables 2 and 3). The comparison of sQTL analysis based on two expression datasets (AtRTD2 and genome assemblies) showed significant overlapping of sQTLs (6181) between two strategies (Supplementary Figure 1). The number of cognate genes increased for the genomic dataset, possibly due to the presence of novel genes/transcripts coming from known/novel genes. The number of transcripts present in the transcriptomic expression dataset is less, but these transcripts were experimentally validated in pilot studies so we used AtRTD2 transcriptomic expression dataset for further analysis. Although the sQTLs were randomly distributed across the genome, 1775 (~28%) of sQTLs localized on chromosome one, whereas chromosome two had the lowest number (956; ~15%) of SNPs linked to splicing patterns (Table 1 and Figure 1). The higher distribution of sQTLs on chromosome one is probably due to its bigger size as compared with other chromosomes (Rhee et al., 2003). To get a better understanding of the influence of the genetic variants (SNPs) on splicing patterns, sQTLs that were within 4 kb from their cognate gene were designated as *cis*-sQTL and every other sQTL outside this window, including those on a different chromosome, as *trans*-sQTLs. Subsequently, 356 *cis*-sQTLs (5% of the total mapped sQTLs) that were associated with the splicing of 301 genes and an extensively high frequency (95%) of *trans*-sQTLs were identified (Table 1 and Figure 2). Interestingly, an overrepresentation of *trans*-sQTLs (*trans*-sQTL hotspots) on chromosome one was observed, which indicates that the molecular factor(s) on this chromosome that may regulates the splicing of several transcripts are *trans*. The sQTLs were then mapped with a list of trait-associated SNPs available in GWAS catalog, and the exact match was found for 911 non-redundant SNPs associated with 757 genes (Supplementary Table 4). Among 911 sQTLs that overlapped with GWAS catalog SNPs, 61 are *cis*-sQTL and are associated with 48 different gene, while 850 are *trans*-sQTL and linked with 709 genes.

To complement the above analysis, we estimated the splicing ratios and AS categories of significant sQTL cognate genes (6129) containing 34,351 transcripts and 26968 AS events (Supplementary Datasets 2 and 3). The association of one of the top sQTL SNP that resides on chromosome 1 at position

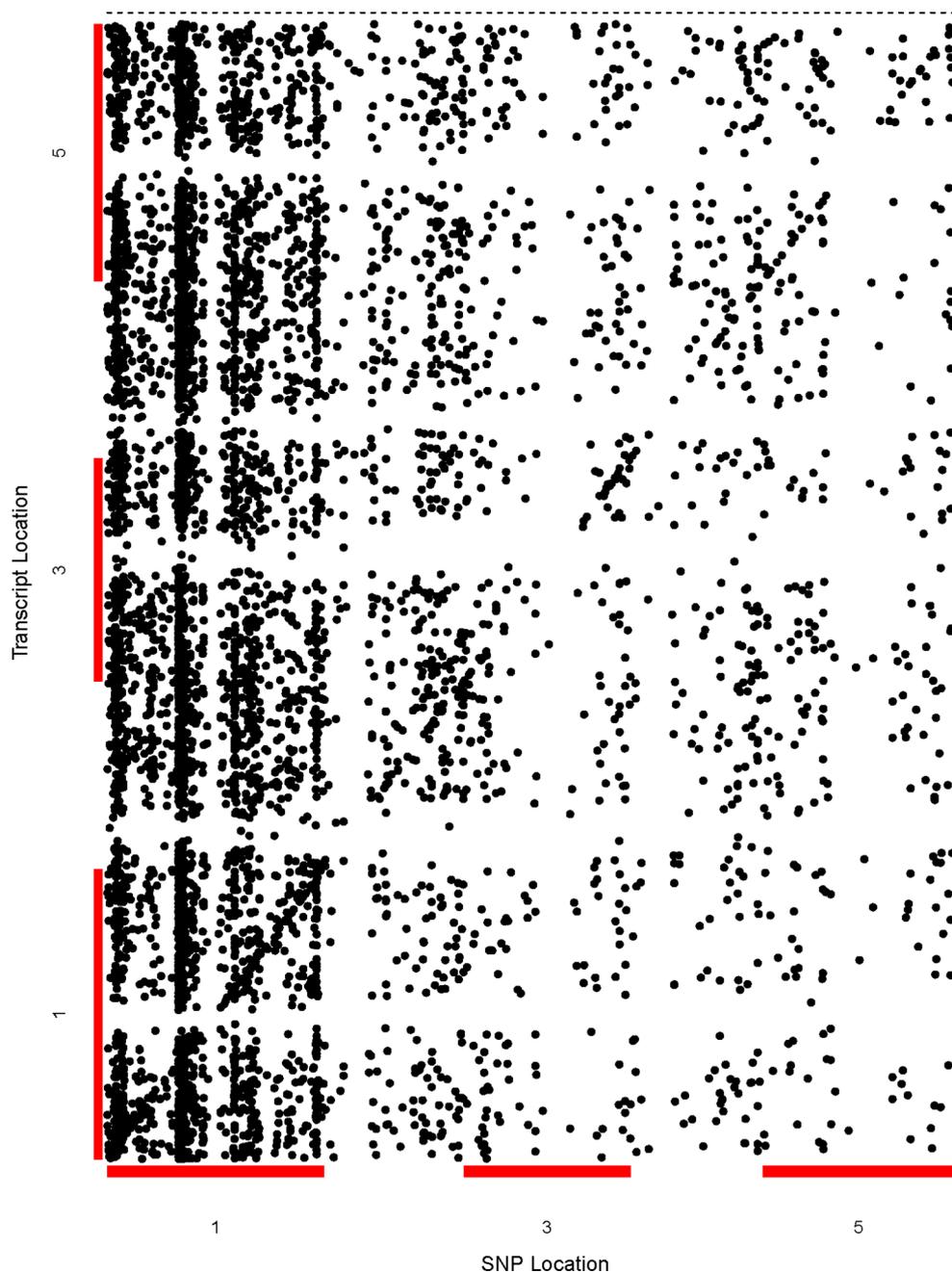
1099063 with splicing ratios of *AT1G04170* (*EIF2 GAMMA; EUKARYOTIC TRANSLATION INITIATION FACTOR 2 GAMMA SUBUNIT*) showed that alteration in genotype from homozygote CC to heterozygote CT significantly modulates splicing ratios of the isoforms *AT1G04170\_JC4* and *AT1G04170\_P1* (Figure 3). Further analysis based on Percent Spliced in (PSI) values highlights the impact of sQTL at AS event level that reflects a significant change in A3 events (Figure 3). The overall AS estimates of sQTL cognate genes shows that IR (10630) was the dominant and mutually exclusive exons (MX) (60) was the least frequent alternative splicing event in diverse accessions of *A. thaliana* (Figure 4). Although less than IR events, the number of alternative 3' splice site (A3) events (8132) was significantly higher than the alternative 5' splice site (A5) events (4536). The number of Alternative first exon (AF) (1827) was close to skipped exons (ES) (1601) but way higher than alternative last exon (AL) (182). The AS events (26968) were subset into *cis* (1590) and *trans* (25378) sQTL cognate gene categories and observed a similar pattern of splicing events. Furthermore, predominance for IR and lower frequency for MX was also observed when analyzed AS genes that emerge as a result of the overlapping of sQTL with GWAS. We also highlighted the subclass of IR events known as Exitrons that are present in coding exons and possess the exonic potential to significantly modulate proteome diversity (Marquez et al., 2015). To illuminate their contribution towards IR events, we extracted exitrons (exon-like introns) by overlapping a list of publicly available 2459 exitrons with 3798 genes possessing 10630 IR event and found 913 common genes (Supplementary Table 5). This AS analysis based on sQTL cognate genes revealed a significant role of IR events in shaping transcriptome diversity and may influence a plant adaptation to complex environmental conditions.

### sQTLs Are Enriched in Exonic Regions

In order to understand the biological significance of our results, we performed functional categorization of filtered SNPs (12,617,361) by classifying them into sQTLs (6,406) and non-sQTLs (12,610,954) SNPs. We were interested in better understanding their role in genome regulation, so we looked at the genomic distribution of sQTL/non-sQTL SNPs and characterized them as exonic SNPs and non-exonic SNPs. The sQTLs were enriched in exonic regions, which show their immense potential to modulate genomic architecture to generate phenotypic diversity. Among exonic variants, missense gene variants showed the highest frequency (1,883) followed by synonymous (1,764) and upstream gene (1,225) variant categories (Figure 5). Among non-exonic regions,

**TABLE 1** | Genome-wide summary of sQTL mapped using 666 diverse ecotypes of *Arabidopsis thaliana*.

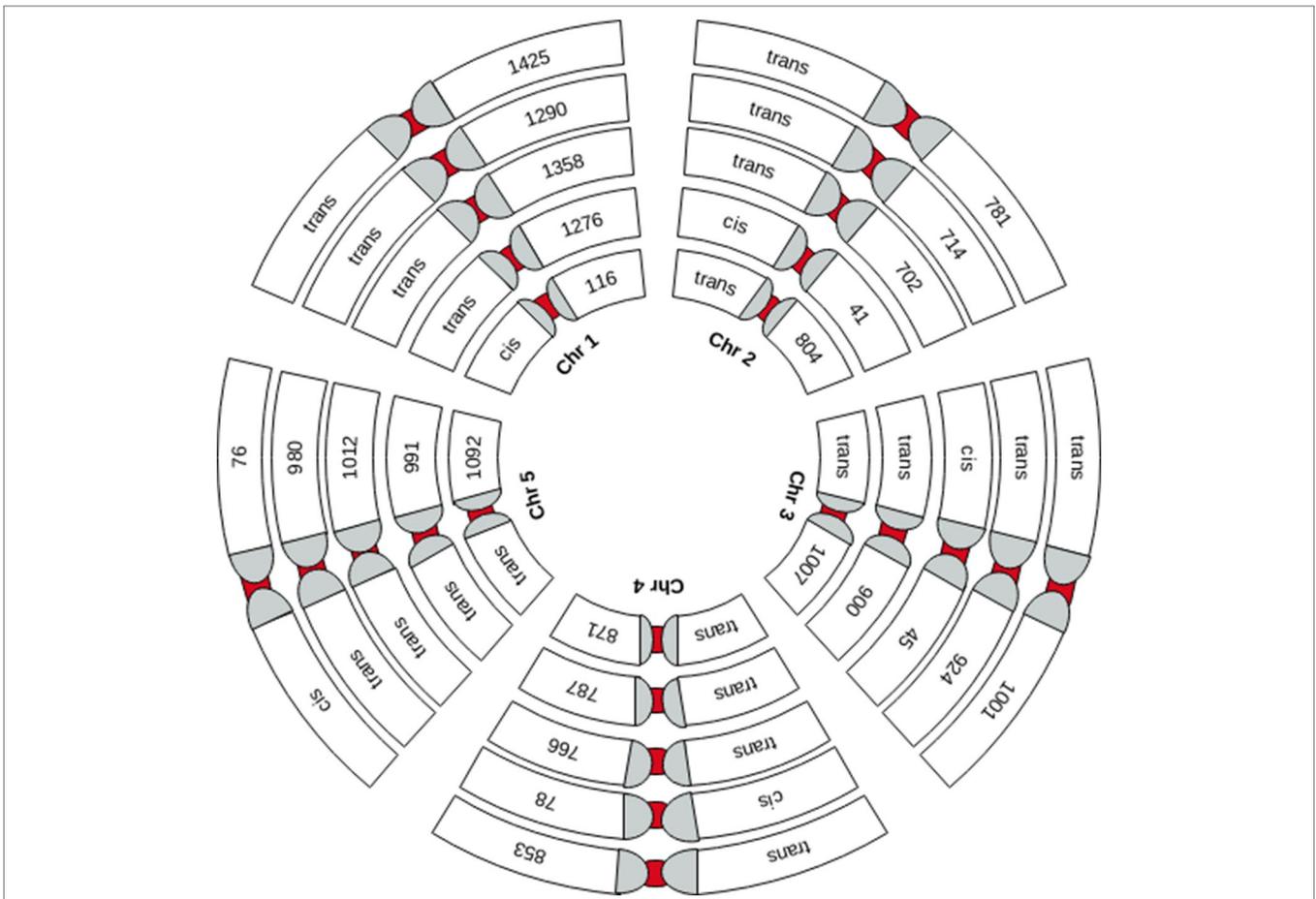
Chromosome	Total sQTLs	Cis-sQTL	Cis-sQTL linked genes	Trans-sQTL	Trans-sQTL linked genes
Chr1	1,775	116	93	5,177	1,741
Chr2	956	41	35	4,688	1,111
Chr3	1,266	45	46	4,773	1,083
Chr4	1,055	78	60	4,641	1,052
Chr5	1,354	76	76	5,112	1,537



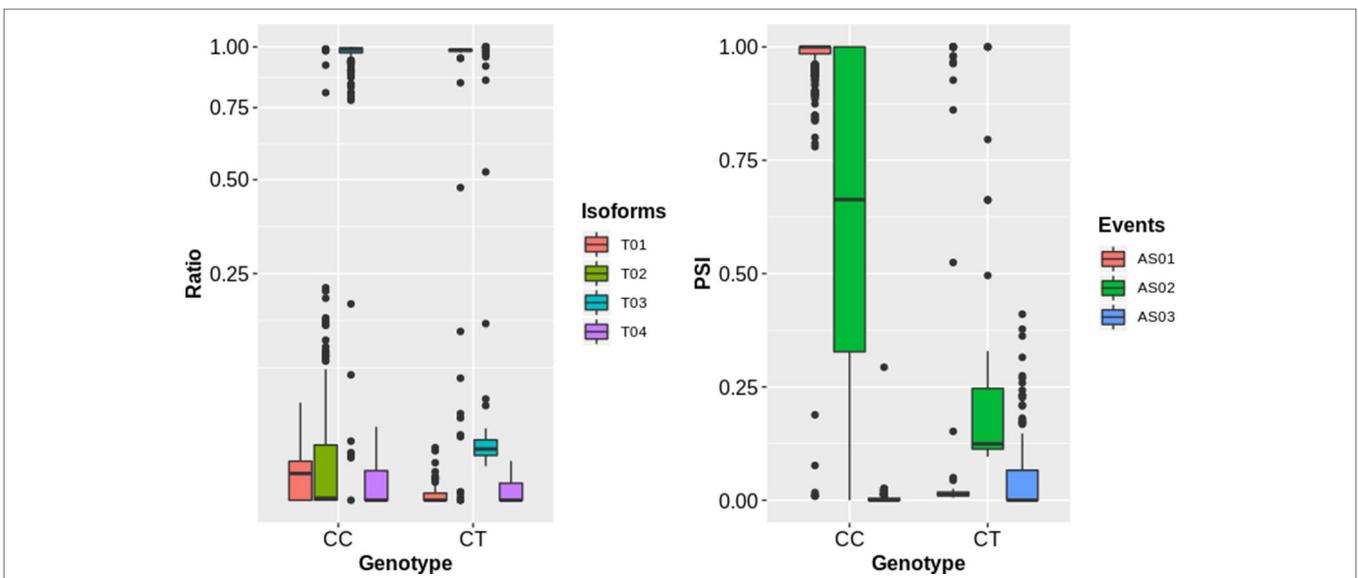
**FIGURE 1 |** Genetic analysis shows the genome-wide distribution of sQTLs and enrichment of *trans*-sQTL on chromosome 1. A comprehensive two-dimensional genome-wide sQTL map showing the distribution of all significant signals for unique sQTLs. Each dot represents an association between a genetic variant location (x-axis) and cognate transcript location (y-axis). The points along the diagonal correspond to *cis*-associations (window 4 kb) with the vertical band representing *trans*-sQTLs.

sQTLs that reside in the intronic regions were significantly higher (282) than intergenic regions (7). Although the non-sQTLs were also enriched in exonic regions, yet within exonic regions, they painted a different picture as compared to sQTLs as they showed a higher number of upstream gene variant regions, compared with the rest of exonic non-sQTLs. However, the proportion of both categories (intron, intergenic) of non-exonic non-sQTLs is almost

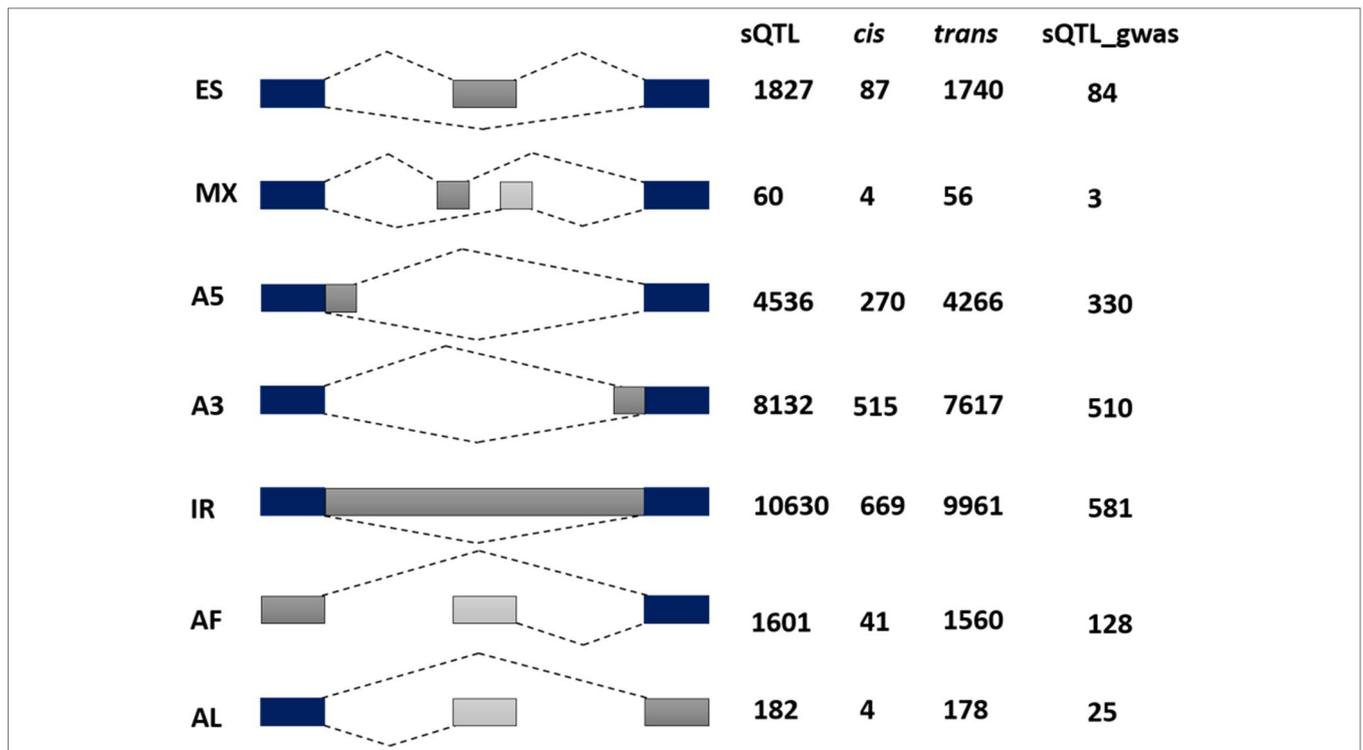
similar. Analysis of the functional context of sQTLs provided by SnpEff revealed a high impact on splice acceptor and stop gained functional categories, although their occurrence in sQTLs is low. The summary of all exonic and non-exonic SNPs for both functional classes (sQTLs and non-sQTLs) showed around 95% SNPs are enriched in variants belong to exonic regions and the majority of them represents the *trans* category, while only 5%



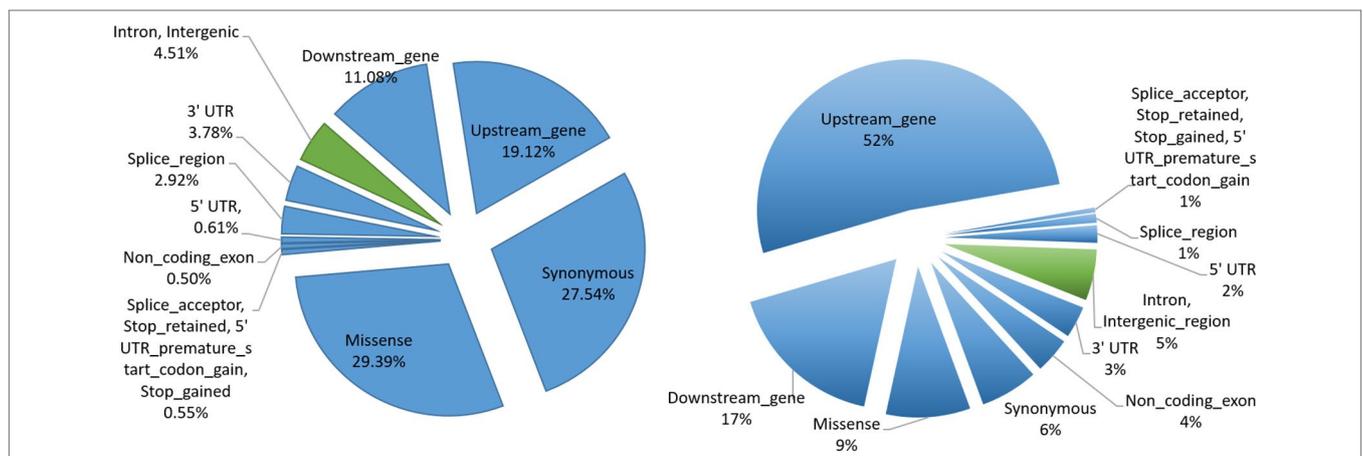
**FIGURE 2 |** Graphical representation of the chromosome-wide distribution of sQTLs. Each group of five chromosomes (Chr1 to 5 is depicted by small to bigger circles, respectively) shows *cis*-sQTLs on one of the five chromosomes and *trans*-sQTLs on other chromosomes.



**FIGURE 3 |** The impact of sQTLs (snp\_1\_1099063) on splicing isoform ratios and AS events of the gene (AT1G04170). The left panel shows the impact of change in genotype on splicing ratios of transcripts (T01: AT1G04170\_c1, T02: AT1G04170\_Jc4, T03: AT1G04170\_P1, T04: AT1G04170\_P2) with sharp change for T01 and T03. The right panel shows the splicing events (AS01, AT1G04170;A3:Chr1:1097120-1097399:1097120-1097405:+, AS02, AT1G04170;A3:Chr1:1097286-1097396:1097286-1097399:+, AS03, AT1G04170;A5:Chr1:1097286-1097399:1097120-1097399:+) with significant change for AS01 and AS02 event.



**FIGURE 4 |** Alternative splicing (AS) categorization of 6129 significant sQTL cognate genes possessing 26968 AS events. Among these 1590 are associated with *cis* and 25378 with the *trans* category. The last column reflects AS pattern of 757 sQTL-GWAS cognate genes possessing 1661 AS events. Among AS categories (exon skipping (ES), mutually exclusive exons (MX), Alternative 5'/3' splice-site (A5/A3), Intron retention (IR), Alternative First/Last Exons (AF/AL), IR are the most common, whereas mutually exclusive exons were the least frequent type of alternative splicing.

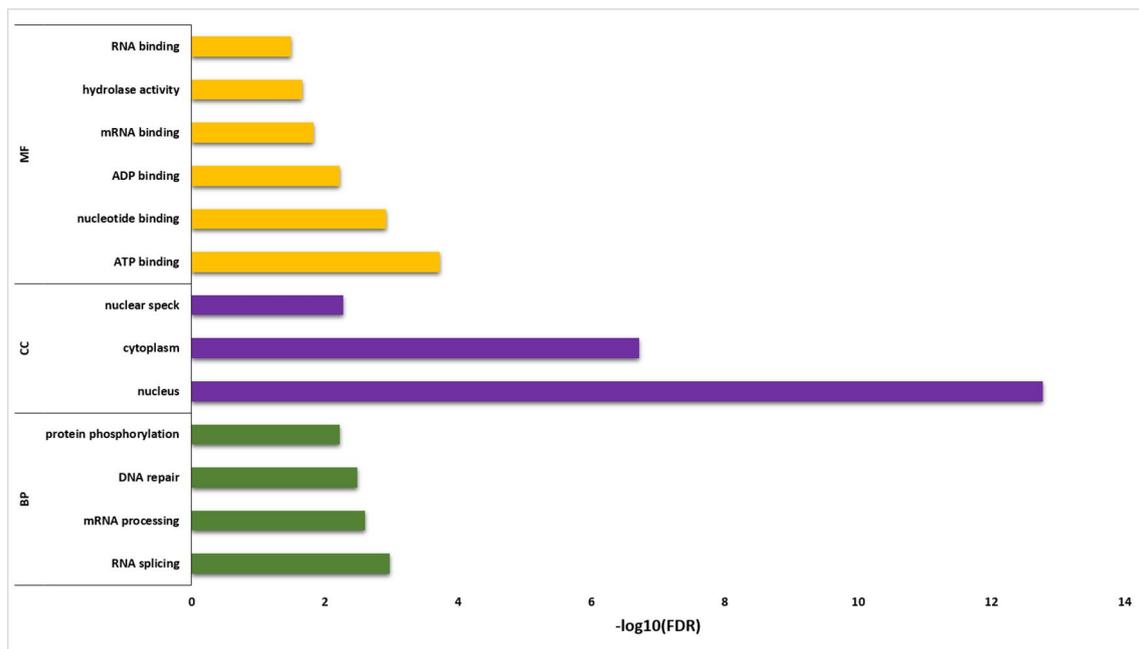


**FIGURE 5 |** sQTL functional characterization. Pie charts indicating proportions of SNPs annotated with each functional category. SNPs in exonic and non-exonic regions are indicated by bluish and greenish colors, respectively. The left shows the functional categorization of sQTL SNPs, whereas the right panel depicts the functional categorization of non-sQTL SNPs.

were found in intronic regions. Furthermore, we have analysed the distribution of *cis* and *trans*-sQTLs along with the exonic locations. Of the 356 *cis*-sQTLs, significantly more (342) reside in exonic locations (hypergeometric *p*-value  $\leq 0.05$ ). Similarly, significantly more (5,775 out of 6,050) *trans*-sQTLs localized in exonic regions (hypergeometric *p*-value  $\leq 0.05$ ).

### Biological Significance of sQTL Cognate Genes

Functional annotation analysis of 6,129 sQTL cognate genes was performed using DAVID (Huang et al., 2009) (**Figure 6; Supplementary Table 6**). The statistically significant gene enrichment terms were filtered based on FDR  $\leq 0.05$  to illuminate



**FIGURE 6 |** Gene enrichment analysis for sQTL-associated genes. Y-axis shows the gene enrichment categories, whereas X-axis illustrates  $-\log_{10}(\text{FDR})$ . BP, CC, and MF represent a biological process, cellular component, and molecular function, respectively.

the momentous role of sQTL cognate genes in diverse biological processes, cellular components, and molecular function. The involvement of sQTL cognate genes in complex biological processes like RNA splicing/processing shows its tremendous potential to modulate the transcriptomic architecture and ability of sQTLs to affect genes associated with DNA repair speculates its critical involvement in genome stability. Furthermore, sQTL association with post-translational modifications in the shape of protein phosphorylation implies their proteome regulatory role that can generate phenotypic diversity.

The presence of sQTL cognate genes in vital cellular components (nucleus, cytoplasm, nuclear speck) and its association with significant binding (RNA, mRNA, ATP, ADP) and catalysis (hydrolase activity) molecular functions highlights its involvement in integral cellular processing that can help in a plant adaptation to the microclimatic conditions.

## Genome Regulatory Role of sQTL

sQTL enrichment within annotated genome regulatory regions was analyzed as mapped sQTLs were spread across the genome and can be enriched among various genome regulatory elements. Moreover, transcription factors (TFs) are important genome regulators as they can mediate transcription by binding in the upstream region of their target genes (Jin et al., 2017). Therefore, the list of TFs from the Plant Transcription Factor Database (PlantTFDB) v4.0 (Jin et al., 2017) was downloaded and showed that sQTLs for 389 genes overlapped with 51 TF families.

Furthermore, the binding of regulatory proteins to *cis*-regulatory DNA elements (CREs) can orchestrate gene expression.

DNase I hypersensitive sites (DHSs) are significantly enriched in CREs that provide chromatin accessibility to regulatory proteins. The DHS sites and nucleosome positioning/occupancy for Arabidopsis were downloaded from PlantDHS database (Zhang et al., 2016) and revealed a significant overlap of sQTLs with CRE enriched regions. The leaf and flower tissue nucleosome positioning data of Arabidopsis was used to see the frequency of sQTLs that reside in nucleosome enriched regions and revealed that 462 sQTLs are flowering specific, 399 are leaf specific, and 4,962 are shared between both tissues. The list of 395 *A. thaliana* splicing-related genes from SRGD (splicing-related gene database) was downloaded to interrogate any overlap with non-redundant sQTL cognate genes and found 128 (7 *cis*- and 121 *trans*-associated) overlapping with splicing-related genes (Supplementary Figure 2; Supplementary Table 7). Among these splicing-related genes (128), the highest number was found on chromosome one (31) and least number on chromosome two (18), which also relates to chromosome size. The overlapped splicing-related genes (128) were associated with 2,397 sQTLs (~37% of overall sQTLs), which show their tremendous potential to serve as significant genome regulatory elements.

## sQTLs Are Enriched Among Stress Responsive, Clock, and Flowering Genes

In order to understand the underlying dynamics of sQTLs with the spatial distribution of 666 accessions worldwide, we analyzed three highly significant gene functional categories (stress response, flowering, and circadian clock) among sQTL associated genes. The three categories are intimately associated with each other and confer

adaptation to different climatic regions. Towards this goal, a list of 3150 Arabidopsis stress-responsive genes was downloaded from STIFDB2 database (Naika et al., 2013). In total, 742 stress-responsive genes associated with significant sQTLs were identified, highlighting the potential of AS in the plant stress response mechanism. Next, we downloaded a list of 346 flowering genes (306 flowering time and 46 flowering development genes) from FLOR-ID database (Bouché et al., 2016) and identified 122 genes that were associated with sQTLs. Similarly, out of 28 core clock genes, 16 were found to be associated with sQTLs (**Supplementary Figure 3A**).

Interestingly, we found six common genes (**Table 2; Supplementary Figure 3B**) between the three groups (circadian, flowering, and stress) that were associated with sQTLs. Besides core clock components like *circadian clock-associated 1* (*CCA1*), *late elongated hypocotyl* (*LHY*), *timing of cab expression 1* (*TOC1*), and pseudo response regulator 7 (*PRR7*), we found *phytochrome interacting factor 5* (*PIF5*) and *b-box domain protein 19* (*BBX19*) genes, which are associated with light-sensing, flowering, and photomorphogenesis, respectively (Suárez-López et al., 2001; Hayama and Coupland, 2003; Somers et al., 2004; Wang et al., 2014; Greenham and McClung, 2015; Wang et al., 2015; Wang and Dehesh, 2015; Nasim et al., 2017; ). Since circadian clock and *PIFs* (*PIF 4/5* and others) influence the timing of flowering by modulating the expression of flowering-related genes, all accessions were divided on the basis of the timing of flowering at 10°C and 16°C (**Supplementary Figure 4**). Clock genes not only control global transcription patterns but their transcript abundance is also modulated by AS (James et al., 2012). Therefore, the mean expression of the aforementioned six genes revealed whether the expression of these genes is associated with the flowering time at 10°C and 16°C. Interestingly, expression of *LHY*, *CCA1*, and *PIF5* is significantly higher for plants growing at 16°C and is accompanied by the low expression of *TOC1* and *BBX19* (flowering repressor) among accessions that flower between 51 and 60 days. The relationship between the expression of these genes and flowering between 61 and 110 days is not straightforward; however, the expression of *PIF5* tremendously increased among accessions that flower late and is accompanied by a lower level of the expression of *BXX19*, which represses flowering by sequestering the *Flowering Time* (*FT*) gene (Wang et al., 2014). On the contrary, the expression of *PIF5* is generally higher among accessions flowering between 51 and 110 days at 16°C; however, late-flowering (131–160 days) accessions show a dramatic increase in the expression of *LHY*. Since most of the late flowering accessions are from Sweden, photoperiod-dependent flowering regulation *via LHY* is more pronounced (Park et al., 2016). Overall, these results indicate that clock, *PIF5*,

and *BXX19* gene mediated flowering time is important for diverse accessions to occupy different geographical regions and many sQTLs mediate these responses.

## DISCUSSION

In this study, we analyzed population-scale transcriptomic and genotypic data of highly diverse 666 *A. thaliana* accessions to comprehensively identify the genomic regions regulating splicing (sQTLs). We used the AtRTD2 database as well as our own genomic assemblies to map sQTLs and found a significant overlap between the two approaches. Since there is significant overlap and the AtRTD2 database is highly validated and non-redundant, we suggest following the transcriptomic approach for sQTL mapping in Arabidopsis and other species where accurate datasets are available. Furthermore, using available transcriptomic datasets and Salmon based approaches are rapid and give reliable results without creating own transcriptome assemblies against the genome. The sQTLs based on the transcriptomic approach are spread genome-wide; however, their frequency varies across chromosomes. The chromosomal distribution relative to the genomic location of their cognate genes (*cis* or *trans*) showed that chromosome one harbors the highest number of *trans*-sQTL hotspots. The chromosomal distribution relative to the genomic location of their cognate genes (*cis* or *trans*) showed that chromosome one harbors the highest number of *trans*-sQTL hotspots. Among the top five associations, two of them reside on chromosome one and belongs to *cis* (snp\_1\_1099063; gene *AT1G04170*) and *trans* (snp\_1\_10688832; gene *AT1G30320*) categories; this strengthens our observation and reflects the immense potential of sQTLs to contribute towards genome complexity and proteome diversity. In order to gain more insight into the biological role and genetic control of splicing, we co-localized the sQTLs with Arabidopsis GWAS hits to illuminate their association with different trait-associated loci and their potential to modulate plants phenotypic variability (Yoo et al., 2016). Functional analysis of sQTLs showed that these genetic variants are significantly localized within exonic regions. Among the exonic regions variants, we discovered a high proportion of missense variants, which can modulate the structure and function of proteins. However, functional annotation painted a different picture by showing the moderate effect of missense variants but the high impact of stop gained and splice acceptor variants, which can modulate AS patterns and proteome diversity (Wachsman et al., 2017).

Although all classes of AS contributed to the sQTLs, intron retention (IR) was prevalent than any other AS type. This is consistent with the previous observations of IR as the most common class of AS and a well-established mechanism for regulating gene expression in plants (Syed et al., 2012; Reddy et al., 2013). However, how IR contribute and/or modulates proteome complexity and the extent to which IR sQTLs influence phenotypic variability remains obscure (Chaudhary et al., 2019). Although IR is associated with nonsense-mediated decay (NMD), this is by no means the only consequence as transcripts can still escape NMD *via* sequestration in the nucleus and some are preferentially recruited to ribosomes (Palusa and Reddy, 2010; Kalyna et al., 2012; Marquez et al., 2012). While our analysis showed that the majority of sQTLs are associated

**TABLE 2** | Association of six genes with sQTL-SNPs and GWAS-SNPs.

Genes ID	Gene name	Association with sQTL	Association with GWAS SNPs
AT5G02810	PRR7	344	18
AT3G59060	PIF5	44	10
AT1G01060	LHY	38	18
AT4G38960	BBX19	38	10
AT2G46830	CCA1	2	13
AT5G61380	TOC1	2	5

with IR events, most SNPs were localized within exonic regions. This is conceivable because only 356 *cis* sQTLs fall in this category and may have a subtle effect on IR in the *cis*-regulatory context. It is plausible that exonic variants cause changes in the pre-mRNA secondary structure, which impacts spliceosome recognition of exon–exon junctions (McManus and Graveley, 2011).

We interrogated the relationship between sQTL associated genes and annotated genome regulatory proteins and found a number of TF families linked to sQTLs. Although experimental validation is required to understand the regulation of TFs *via* AS, these results elaborate the potential of splicing as a mechanism for regulating gene expression (Nasim et al., 2017; Takata et al., 2017). Similarly, we also found a strong overlap between sQTLs, CREs, and nucleosome occupancy, especially among flower and leaf specific genes. Since nucleosome occupancy is much higher in exons, it helps to define intron–exon definition (Schwartz et al., 2009) to orchestrate an appropriate splicing response under variable environmental conditions. In addition, nucleosome occupancy has a strong influence on RNA polymerase II processing as its speed tends to be higher in regions with more open chromatin structure and Pol II speed regulates AS (Ullah et al., 2018; Godoy Herz et al., 2019). It is therefore not surprising that many sQTLs are enriched in the DHS site and have a strong association with regions with higher nucleosome occupancy among important life-history traits like leaf and flower. This data also indicates that many of the important traits like flowering are regulated *via* epigenetic means, and this is reminiscent of the downregulation of the *flowering* repressor locus *c* (*FLC*) that regulates flowering in a cold-dependent manner (Michaels and Amasino, 1999; Shindo et al., 2006).

Functional characterization of sQTLs among stress-responsive, circadian clock, and flowering genes revealed six common genes (*CCA1*, *LHY*, *TOC1*, *PRR7*, *PIF5*, and *BXX19*) that are shared in these categories (**Supplementary Figure 5, Supplementary Table 8**). Among these six genes, *PRR7* (*AT5G02810*) showed high significance in the sQTL analysis by accruing second rank (**Supplementary Dataset 2**). *PRR7* was impacted by the transQTL (*snp\_5\_626009*) present on chromosome 5 at position 626009, and this SNP modulated the splicing ratios and affected the AS patterns. We are aware that co-localisation of sQTLs in these categories would need further validation but we hope that it would provide an interesting starting point and a list of useful genes that may be regulated *via* alternative splicing. It is well known that the circadian clock plays a vital role in the normal functioning of plants and is intimately associated with carbon fixation during the day and starch mobilization to promote growth during the night (Dodd et al., 2005; Graf et al., 2010). Rhythmicity of clock components is not only associated with appropriate growth responses under variable and often stressful conditions but also promotes fitness and adaptive responses in plants (Dodd et al., 2005). Interestingly, down-regulation of *LHY* and *CCA1* around noon time among Arabidopsis hybrids and polyploids promotes heterosis *via* upregulation of chlorophyll, starch synthesis, and metabolism genes (Ni et al., 2009). Intriguingly, *1-aminocyclopropane-1-carboxylate synthase* (*ACS*, a rate-limiting enzyme in ethylene synthesis) is also downregulated by *CCA1* and phytochrome-interacting factor 5 (*PIF5*) during the

day and night, respectively, to promote heterosis in Arabidopsis (Song et al., 2018). Further, *LHY* and *CCA1* have been implicated in cold temperature acclimation responses and may also be important under higher temperature and stressful temperature conditions. *CCA1* and *LHY* are partially redundant but show different expression and splicing patterns under cold conditions (Calixto et al., 2018). Similarly, *PRR7* has been associated with temperature responses in Arabidopsis and also shows different splicing patterns under normal and cold conditions (James et al., 2012). Recent experimental and modeling data suggest that *TOC1*, in concert with *ABA* levels, plays the role of an environmental sensor that coordinates the pace of the central oscillator to affect downstream processes (Pokhilko et al., 2013). Mechanistic details of *TOC1* mediated drought responses were revealed by analyzing the relationship between *TOC1* and an *ABA*-related gene (*ABAR/CHLH/GUN5*) (Legnaioli et al., 2009; Castells et al., 2010). Under drought conditions and elevated *ABA* levels, *TOC1* binds the *ABAR* promoter and modulates its circadian expression, resulting in clock-dependent gating of *ABA* function and drought tolerance (Legnaioli et al., 2009). Recent findings revealed that *LHY* modulates the expression of many genes involved in *ABA* signaling pathway to fine-tune plant performance under drought and osmotic stress conditions. However, *LHY* also maintained seed germination and plant growth *via* alleviation of the inhibitory effect of *ABA* (Adams et al., 2018).

The timing of flowering is crucial for the survival and adaptation of diverse Arabidopsis accessions in different geographical regions of the world. It is well known that the timing of flowering has a huge impact on seed set, grain filling, maturation, and overall productivity. Intriguingly, plants can also speed up their flowering in response to various environmental fluctuations and stresses, presumably as a survival strategy to produce seeds as quickly as possible (Wang and Dehesh, 2015). It is not surprising that one of the common genes between clock, flowering, and stress-related genes is *BBX19*. This gene has been shown to work as a circadian clock output and downregulates *CONSTANS* (*CO*) and precisely times the expression of *flowering locus T* (*FT*) in a day-length dependent manner to orchestrate appropriate flowering time (Wang et al., 2015). Interestingly, E3 ubiquitin ligase *constitutive photomorphogenic 1* (*COP1*), *early flowering 3* (*ELF3*), *phytochrome-interacting factor 4* (*PIF4*), and *PIF5* also influence *BBX19* function to mediate photomorphogenic responses in Arabidopsis (Wang et al., 2015). Furthermore, expression of *BBX19* is significantly reduced as a result of high levels of *methylerythritol cyclodiphosphate* (*MEcPP*), which is a plastidial isoprenoid intermediate that also functions as a stress-response retrograde signal to orchestrate appropriate transcriptional response (Xiao et al., 2012; Wang and Dehesh, 2015). Therefore, the *BBX19* gene provides the role of a flowering checkpoint that links a stress-specific retrograde signal (*MEcPP*) *via* sequestering the active *CO* gene, which is essential for *FT* transcription to promote flowering (Wang and Dehesh, 2015). Higher expression of *BBX19* delays flowering; however, its expression is positively correlated with *PIF5* expression to promote hypocotyl growth. Taken together, *BBX19* plays a dual role to modulate plant growth and stress-responsive flowering (Wang et al., 2015). Therefore, it is not surprising that

many sQTL SNPs are associated with these genes and may be important for defining phenotypes and underlying genotypes among geographically diverse lines of Arabidopsis. Since *BBX19* controls flowering in a clock-regulated and stress-dependent manner, we propose that expression and sQTL patterns of *BBX19* may have a bearing on the flowering patterns of 666 accessions from different geographical regions of the world and play a role in adaptive responses. Recent evidence also shows that SNPs are present at almost every 200 bp in different ecotypes of Arabidopsis and can alter genomic architecture to affect splicing efficiency, gene expression pattern, and phenotypic diversity (Gan et al., 2011). We envisage that the presence of widespread variation in diverse ecotypes of Arabidopsis and our sQTL analysis would provide a solid platform for sQTL discovery and their influence on phenotypic traits in the future.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Genotype data can be found here: <https://1001genomes.org/data/GMI-MPI/releases/v3.1/> and RNA-Seq data can be found here: <http://signal.salk.edu/1001.php>

## AUTHOR CONTRIBUTIONS

NS and WK conceived the study, WK performed most of the analysis, and MH, AR, SC, and IJ contributed to it. All authors contributed toward preparing the manuscript. WK and MH contributed equally.

## ACKNOWLEDGMENTS

We thank the funding agencies for research support: Leverhulme Trust [RPG-2016-014]; DOE Office of Science, Office of Biological and Environmental Research (BER) [DE-SC0010733]; National Science Foundation; and the US Department of Agriculture (ASNR). MH is supported by a University of Edinburgh Chancellor's fellowship and in part by Bill and Melinda Gates Foundation and with UK aid from the UK Government Department for International Development (Grant Agreement OPP1127286)

## REFERENCES

- Adams, S., Grundy, J., Veflingstad, S. R., Dyer, N. P., Hannah, M. A., Ott, S., et al. (2018). Circadian control of abscisic acid biosynthesis and signalling pathways revealed by genome-wide analysis of LHY binding targets. *New Phytol.* 220 (3), 893–907. doi: 10.1111/nph.15415
- Alamancos, G. P., Pagès, A., Trincado, J. L., Bellora, N., and Eyra, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* 21, 1521–1531. doi: 10.1261/rna.051557.115
- Alamancos, G., Pagès, A., Trincado, J., and Bellora, N. (2014). SUPPA: a super-fast pipeline for alternative splicing analysis from RNA-Seq. *bioRxiv* 008763. doi: 10.1101/008763
- Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., et al. (2016). 1,135 Genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell* 166, 481–491. doi: 10.1016/j.cell.2016.05.063

under the auspices of the Centre for Tropical Livestock Genetics and Health, established jointly by the University of Edinburgh, Scotland's Rural College, and the International Livestock Research Institute. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of Bill & Melinda Gates Foundation nor the UK Government. The Roslin Institute receives institute strategic programme funds from Biotechnology and Biological Sciences Research Council (BBSRC).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01160/full#supplementary-material>

**SUPPLEMENTARY FIGURE 1** | Complementarity of the genome and transcriptome-based approaches for sQTL analysis.

**SUPPLEMENTARY FIGURE 2** | Summary of sQTL cognate genes overlapping with splicing related genes. The x-axis shows the type of splicing related genes and Y-axis exhibits the number of input and matched splicing related genes.

**SUPPLEMENTARY FIGURE 3** | Functional characterization of genes associated with sQTL. (A) sQTL cognate genes were highly enriched in different core functional categories (e.g. stress response, flowering). (B) Six genes were shared between different functional categories.

**SUPPLEMENTARY FIGURE 4** | Phenotypic association of six genes with flowering time. The x-axis shows days of flowering at 10 °C (A) and 16 °C (B) and Y-axis show the average/relative gene expression value for six genes across a diverse set of 666 A. thaliana accessions.

**SUPPLEMENTARY FIGURE 5** | The impact of sQTLs on splicing isoform ratios and AS events of six genes. The left panel shows the impact of change in genotype on splicing ratios of transcripts and the right panel shows the splicing events. For a detailed description of all isoforms and transcripts (see **Supplementary Table 8**).

**SUPPLEMENTARY DATASET 1** | RNA-Sequencing Analysis using genome as a reference

**SUPPLEMENTARY DATASET 2** | (Splicing ratios of sQTL cognate Genes): <https://figshare.com/s/4dbd07a41f0fd4c6988c>

**SUPPLEMENTARY DATASET 3** | (Percent Spliced In (PSI) of sQTL cognate Genes): <https://figshare.com/s/89a7123e21d7de9643c3>

**SUPPLEMENTAL DATASET 4** | (Stringtie merged assembly): <https://figshare.com/s/fd1a95af9e2b3025b8ff>

- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Res.* 44, D1167–D1171. doi: 10.1093/nar/gkv1054
- Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., et al. (2018). Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome. *Plant Cell* 30, 1424–1444. doi: 10.1105/tpc.18.00177
- Castells, E., Portolés, S., Huang, W., and Mas, P. (2010). A functional connection between the clock component TOC1 and abscisic acid signaling pathways. *Plant Signal. Behav.* 5 (4), 409–411. doi: 10.4161/psb.5.4.11213
- Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., and Syed, N. H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* xx, 1–11. doi: 10.1016/j.tplants.2019.02.006
- Chen, Q., Han, Y., Liu, H., Wang, X., Sun, J., Zhao, B., et al. (2018). Genome-wide association analyses reveal the importance of alternative splicing in

- diversifying gene function and regulating phenotypic variation in maize. *Plant Cell* 30, 1404–1423. doi: 10.1105/tpc.18.00109
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dodd, A. N., Toth, R., Kevei, E., Webb, A. A. R., Salathia, N., Hibberd, J. M., et al. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309, 630–633. doi: 10.1126/science.1115581
- Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., et al. (2015). SKIP Confers Osmotic Tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. *Mol. Plant* 8, 1038–1052. doi: 10.1016/j.molp.2015.01.011
- Filichkin, S. A., Cumbie, J. S., Dharmawardhana, P., Jaiswal, P., Chang, J. H., Palusa, S. G., et al. (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant* 8, 207–227. doi: 10.1016/j.molp.2014.10.011
- Filichkin, S. A., Hamilton, M., Dharmawardhana, P. D., Singh, S. K., Sullivan, C., Ben-Hur, A., et al. (2018). Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Front. Plant Sci.* 9, 1–17. doi: 10.3389/fpls.2018.00005
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Gen. Res.* 20 (1), 45–58. doi: 10.1101/gr.093302.109
- Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., et al. (2018). A Genomic Map of Climate Adaptation in Arabidopsis thaliana at a Micro-Geographic Scale. *Front. Plant Sci.* 9, 967. doi: 10.3389/fpls.2018.00967
- François, O., Blum, M. G. B., Jakobsson, M., and Rosenberg, N. A. (2008). Demographic history of European populations of Arabidopsis thaliana. *PLoS Genet.* 4 (5), e1000075. doi: 10.1371/journal.pgen.1000075
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., and Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature* 477 (7365), 419–423. doi: 10.1038/nature10414
- Godoy Herz, M. A., Kubaczka, M. G., Brzyżek, G., Servi, L., Krzysztos, M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* 73 (5), 1066–1074.e3. doi: 10.1016/j.molcel.2018.12.005
- Graf, A., Schlereth, A., Stitt, M., and Smith, A. M. (2010). Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proc. Natl. Acad. Sci.* 107 (20), 9458–9463. doi: 10.1073/pnas.0914299107
- Greenham, K., and McClung, C. R. (2015). Integrating circadian dynamics with physiological processes in plants. *Nat. Rev. Genet.* 16(10), 598–610. doi: 10.1038/nrg3976
- Han, S., Jung, H., Lee, K., Kim, H., and Kim, S. (2017). Genome wide discovery of genetic variants affecting alternative splicing patterns in human using bioinformatics method. *Genes Genomics* 39, 453–459. doi: 10.1007/s13258-016-0466-7
- Hartmann, L., Wiefner, T., and Wachter, A. (2018). Subcellular compartmentation of alternatively-spliced transcripts defines SERINE/ARGININE-RICH PROTEIN 30 expression. *Plant Physiol.* 176, 2886–2903. doi: 10.1104/pp.17.01260
- Hayama, R., and Coupland, G. (2003). Shedding light on the circadian clock and the photoperiodic control of flowering. *Curr. Opin. Plant Biol.* 6 (1), 13–19. doi: 10.1016/S1369-5266(02)00011-0
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Huang, J., Gu, L., Zhang, Y., Yan, T., Kong, G., Kong, L., et al. (2017). An oomycete plant pathogen reprograms host pre-mRNA splicing to subvert immunity. *Nat. Commun.* 8, 2051. doi: 10.1038/s41467-017-02233-5
- Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* 23, 1085–1094. doi: 10.1093/molbev/msj118
- Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47 (6), 2716–2726. doi: 10.1093/nar/gkz121
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012). Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell.* 24 (3), 961–981. doi: 10.1105/tpc.111.093948
- Jiang, J., Liu, X., Liu, C., Liu, G., Li, S., and Wang, L. (2017). Integrating omics and alternative splicing reveals insights into grape response to high temperature. *Plant Physiol.* 173, 1502–1518. doi: 10.1104/pp.16.01305
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Kalyna, M., Lopato, S., Voronin, V., and Barta, A. (2006). Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34 (16), 4395–4405. doi: 10.1093/nar/gkl570
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40 (6), 2454–2469. doi: 10.1093/nar/gkr932
- Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urlich, M. A., et al. (2016). Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. *Cell* 166, 492–505. doi: 10.1016/j.cell.2016.06.044
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. doi: 10.1111/tj.13312
- Krueger, F. (2015). Trim galore!: a wrapper tool around cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.
- Lee, A., Lee, S. S., Jung, W. Y., Park, H. J., Lim, B. R., Kim, H. S., et al. (2016). The OsCYP19-4 gene is expressed as multiple alternatively spliced transcripts encoding isoforms with distinct cellular localizations and PPLase activities under cold stress. *Int. J. Mol. Sci.* 17, 1154. doi: 10.3390/ijms17071154
- Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., et al. (2017). On the post-glacial spread of human commensal Arabidopsis thaliana. *Nat. Commun.* 8, 14458. doi: 10.1038/ncomms14458
- Legnaioli, T., Cuevas, J., and Mas, P. (2009). TOC1 functions as a molecular switch connecting the circadian clock with plant responses to drought. *EMBO J.* 28 (23), 3745–3757. doi: 10.1038/emboj.2009.297
- Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., et al. (2013). An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing. *Plant Physiol.* 162, 512–521. doi: 10.1104/pp.112.205864
- Liu, Z., Qin, J., Tian, X., Xu, S., Wang, Y., Li, H., et al. (2018). Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (Triticum aestivum L.). *Plant Biotechnol. J.* 16, 714–726. doi: 10.1111/pbi.12822
- Lutz, U., Posé, D., Pfeifer, M., Gundlach, H., Hagmann, J., Wang, C., et al. (2015). Modulation of ambient temperature-dependent flowering in Arabidopsis thaliana by natural variation of FLOWERING LOCUS M. *PLoS Genet.* 11 (10), e1005588. doi: 10.1371/journal.pgen.1005588
- Marquez, Y., Brown, J. W. S., Simpson, C., and Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22 (6), 1184–1195. doi: 10.1101/gr.134106.111
- Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- McManus, C. J., and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* 21 (4), 373–379. doi: 10.1016/j.gde.2011.04.001
- Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell Online.* 11 (5), 949–956. doi: 10.1105/tpc.11.5.949
- 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. doi: 10.1038/ncomms5698
- Naika, M., Shameer, K., Mathew, O. K., Gowda, R., and Sowdhamini, R. (2013). STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant Cell Physiol.* 54, e8–e8. doi: 10.1093/pcp/pcs185
- Nasim, Z., Fahim, M., and Ahn, J. H. (2017). Possible role of MADS AFFECTING FLOWERING 3 and B-BOX DOMAIN PROTEIN 19 in flowering time

- regulation of Arabidopsis mutants with defects in nonsense-mediated mRNA decay. *Front. Plant Sci.* 8, 191. doi: 10.3389/fpls.2017.00191
- Ni, Z., Kim, E. D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457 (7227), 327–331. doi: 10.1038/nature07523
- Pajoro, A., Severing, E., Angenent, G. C., and Immink, R. G. H. (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.* 18 (1), 102. doi: 10.1186/s13059-017-1235-x
- Palusa, S. G., Ali, G. S., and Reddy, A. S. N. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.* 49, 1091–1107. doi: 10.1111/j.1365-313X.2006.03020.x
- Palusa, S. G., and Reddy, A. S. N. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytol.* 185, 83–89. doi: 10.1111/j.1469-8137.2009.03065.x
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* 102, 11–26. doi: 10.1016/j.ajhg.2017.11.002
- Park, M. J., Kwon, Y. J., Gil, K. E., and Park, C. M. (2016). LATE ELONGATED HYPOCOTYL regulates photoperiodic flowering via the circadian clock in Arabidopsis. *BMC Plant Biol.* 16, 1–12. doi: 10.1186/s12870-016-0810-8
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122
- Pokhilko, A., Mas, P., and Millar, A. J. (2013). Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Syst. Biol.* 7 (1), 7–23. doi: 10.1186/1752-0509-7-23
- Qu, W., Gurdziel, K., Pique-Regi, R., and Ruden, D. M. (2017). Identification of splicing quantitative trait loci (sQTL) in *Drosophila melanogaster* with developmental late (Pb2+) exposure. *Front. Genet.* 8, 145. doi: 10.3389/fgene.2017.00145
- Reddy, A., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25 (10), 3657–3683. doi: 10.1105/tpc.113.117523
- Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., et al. (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 31, 224–228. doi: 10.1093/nar/gkg076
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16 (9), 990–995. doi: 10.1038/nsmb.1659
- Seaton, D. D., Graf, A., Baerenfaller, K., Stitt, M., Millar, A. J., and Gruissem, W. (2018). Photoperiodic control of the Arabidopsis proteome reveals a translational coincidence mechanism. *Mol. Syst. Biol.* 14, e7962. doi: 10.15252/msb.20177962
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response. *Genes Dev.* 20 (22), 3079–3083. doi: 10.1101/gad.405306
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006
- Smith, C. C. R., Tittes, S., Mendieta, J. P., Collier-zans, E., Rowe, H. C., Rieseberg, L. H., et al. (2018). Genetics of alternative splicing evolution during sunflower domestication. *Proc. Natl. Acad. Sci.* 115 (26), 6768–6773. doi: 10.1073/pnas.1803361115
- Somers, D. E., Kim, W.-Y., and Geng, R. (2004). The F-Box protein ZEITLUPE confers dosage-dependent control on the circadian clock, photomorphogenesis, and flowering time. *Plant Cell Online* 16 (3), 769–782. doi: 10.1105/tpc.016808
- Song, Q., Ando, A., Xu, D., Fang, L., Zhang, T., Huq, E., et al. (2018). Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc. Natl. Acad. Sci.* 115 (21), 5606–5611. doi: 10.1073/pnas.1722068115
- Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* 410 (6832), 1116–1120. doi: 10.1038/35074138
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants - coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- 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. doi: 10.1038/ncomms14519
- Togninalli, M., Seren, Ü., Meng, D., Fitz, J., Nordborg, M., Weigel, D., et al. (2018). The AraGWAS Catalog: a curated and standardized Arabidopsis thaliana GWAS catalog. *Nucleic Acids Res.* 46, D1150–D1156. doi: 10.1093/nar/gkx954
- Ullah, F., Hamilton, M., Reddy, A. S. N., and Ben-Hur, A. (2018). Exploring the relationship between intron retention and chromatin accessibility in plants. *BMC Genomics* 19, 21. doi: 10.1186/s12864-017-4393-z
- Wachsman, G., Modliszewski, J. L., Valdes, M., and Benfey, P. N. (2017). A SIMPLE pipeline for mapping point mutations. *Plant Physiol.* 174, 1307–1313. doi: 10.1104/pp.17.00415
- Wang, C.-Q., Guthrie, C., Sarmast, M. K., and Dehesh, K. (2014). BBOX1 interacts with CONSTANS to repress FLOWERING LOCUS T transcription, defining a flowering time checkpoint in Arabidopsis. *Plant Cell* 26 (9), 3589–3602. doi: 10.1105/tpc.114.130252
- Wang, C.-Q., Sarmast, M. K., Jiang, J., and Dehesh, K. (2015). The transcriptional regulator BBOX1 promotes hypocotyl growth by facilitating COP1-mediated EARLY FLOWERING3 degradation in Arabidopsis. *Plant Cell*. doi: 10.1105/tpc.15.00044
- Wang, C., and Dehesh, K. (2015). From retrograde signaling to flowering time. *Plant Signal. Behav.* 10(6), e1022012. doi: 10.1080/15592324.2015.1022012
- Xiao, Y., Savchenko, T., Baidoo, E. E. K., Chehab, W. E., Hayden, D. M., Tolstikov, V., et al. (2012). Retrograde signaling by the plastidial metabolite MeCPP regulates expression of nuclear stress-response genes. *Cell* 149 (7), 1525–1535. doi: 10.1016/j.cell.2012.04.038
- Yang, Q., Hu, Y., Li, J., and Zhang, X. (2017). ulfasQTL: an ultra-fast method of composite splicing QTL analysis. *BMC Genomics* 18, 963. doi: 10.1186/s12864-016-3258-1
- Yoo, W., Kyung, S., Han, S., and Kim, S. (2016). Investigation of splicing quantitative trait loci in *Arabidopsis thaliana*. *Genomics Inform.* 14, 211–215. doi: 10.5808/GI.2016.14.4.211
- Zhang, P., Deng, H., Xiao, F. M., and Liu, Y. S. (2013). Alterations of alternative splicing patterns of Ser/Arg-Rich (SR) genes in response to hormones and stresses treatments in different ecotypes of rice (*Oryza sativa*). *J. Integr. Agric.* 12, 737–748. doi: 10.1016/S2095-3119(13)60260-9
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267
- Zhang, R., Calixto, C. P. G., Tzioutziou, N. A., James, A. B., Simpson, C. G., Guo, W., et al. (2015). AtRTD - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in *Arabidopsis thaliana*. *New Phytol.* 208, 96–101. doi: 10.1111/nph.13545
- Zhang, T., Marand, A. P., and Jiang, J. (2016). PlantDHS: a database for DNase I hypersensitive sites in plants. *Nucleic Acids Res.* 44 (D1), D1148–53. doi: 10.1093/nar/gkv962
- Zhang, X., Cal, A. J., and Borevitz, J. O. (2011). Genetic architecture of regulatory variation in *Arabidopsis thaliana*. *Genome Res.* 21, 725–733. doi: 10.1101/gr.115337.110

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Khokhar, Hassan, Reddy, Chaudhary, Jabre, Byrne and Syed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.