Perspective on alternative splicing and proteome complexity in plants

Saurabh Chaudhary[1*], Ibtissam Jabre[1*], Anireddy SN Reddy[2], Dorothee Staiger[3] and Naeem H Syed[1]

1 School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, CT1 1QU, UK

2 Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO 80523-1878 USA.

3 RNA Biology and Molecular Physiology, Faculty of Biology, Bielefeld University, Bielefeld, Germany.

*Equal Contribution

Correspondence: naeem.syed@canterbury.ac.uk (N.H. Syed)

KEYWORDS Alternative Splicing, IDPs, Protein Diversity, Stress Memory, Translational Coincidence

ABSTRACT

Alternative Splicing (AS) generates multiple transcripts from the same gene, however AS contribution to proteome complexity remains elusive in plants. AS is prevalent under stress conditions in plants, but it is counterintuitive why plants would invest in protein synthesis under declining energy supply. We propose that plants employ AS not only to potentially increasing proteomic complexity, but also to buffer against the stress-responsive transcriptome to reduce the metabolic cost of translating all AS transcripts. To maximise efficiency under stress, plants may make fewer proteins with disordered domains via AS to diversify substrate specificity and maintain sufficient regulatory capacity. Furthermore, we suggest that chromatin state-dependent AS engenders short/long-term stress memory to mediate reproducible transcriptional response in the future.

Regulation of Proteome Complexity by Alternative Splicing

As sessile organisms, plants exert a tight control over their gene expression patterns under normal and stress conditions to maximise carbon fixation and resource allocation efficiency to promote growth and fitness in the short and long term [1]. AS adds another layer of complexity to modulate transcriptome diversity [2–4] and potentially proteome complexity in a tissue- and condition-dependent manner [5,6]. It is well established that AS often allows fine-tuning of gene expression by changing the ratios of productive and unproductive variants [7,8]. However, limited data is available on the contribution of AS to protein diversity in plants [5]. Recent transcriptome and translatome data from humans suggest a significant contribution of AS towards protein diversity [9–14]. However, relatively few alternative isoforms have been discovered in various proteomic studies that encode different proteins [15–19]. The scientific community is divided on this issue and some argue that poor sensitivity of **Mass-Spectrometry (MS)** techniques is a major limitation to detect changes in protein isoforms as a result of AS (**Box 1**) [18]. On the other hand, it is also proposed that not all alternative isoforms are biologically important, because alternative transcripts are generally a recent evolutionary innovation and under neutral selection [17]. Since limited proteomic data is available in plants, it is paramount to perform comprehensive proteomic studies in different tissues and in response to diverse stresses to illuminate the contribution of AS towards protein diversity and/or increasing regulatory capacity in plants. In addition, global analysis of translation patterns of splice isoforms needs to be studied in different tissues and stresses at multiple time points throughout the diurnal cycle.

Transcription and translation are energetically expensive [20], nonetheless plants exhibit a higher level of AS under stressful conditions [21]. This scenario poses potential problems, for example, if the aim is to diversify the proteome then why plants should invest

in translation when photosynthetic capacity declines in stress conditions? Moreover, AS frequently generates transcripts harbouring premature termination codon (PTC+), which are degraded by the nonsense-mediated decay (NMD) pathway [22–24]. NMD is a cytoplasmic mRNA quality control mechanism that targets newly synthesised capped transcripts harbouring NMD+ features during the pioneer round of translation [25,26]. Interestingly, evidence from humans suggests that NMD is not restricted to the pioneer round of translation and could also be triggered for already translating mRNAs as a result of change in the cellular environment and/or needs [27,28]. Among all AS events, **intron retention** (IR) is the most prevalent event in plants [22,23]. Most IR transcripts are predominantly sequestered in the nucleus under a particular stress or developmental stage for further processing upon cell requirement or degraded by the NMD pathway [29–32]. Some IR transcripts carry introns with features of protein-coding exons, which are termed as exitrons, and splicing of these exitrons affects protein functionality [6,33]. Exitrons and other types of splice variants can often lead to the formation of **Intrinsically disordered proteins or regions** (IDPs/IDRs) [6,34]. IDPs and IDRs lack fixed three-dimensional structure due to their amino acid composition, which prevents appropriate hydrophobic region formation [35]. Importantly, variation in the three-dimensional structure of proteins, as a result of AS and **post-translational modifications** (PTMs), results in the diversification of substrate specificity and enhanced regulatory capacity [36–39].

Although AS coupled to NMD plays a major role in regulating the Arabidopsis (*Arabidopsis thaliana*) transcriptome [40] and potentially protein levels, however, most of the PTC+ transcripts (IR and others) if translated, would produce truncated proteins **(Figure 1)** and create a very toxic environment to carry out the normal activity of the cell [41]. The efficiency of NMD during and after the pioneer round of translation is robust and most PTC+ transcripts are rapidly degraded upon their arrival in the cytoplasm [27,28,42]. Intriguingly,

NMD responses are dampened in both mammals and plants under stress conditions and this strategy may facilitate an appropriate response via translating some of the stress-responsive genes and splice variants [43]. We propose that under initial episodes of stress conditions, plants buffer against normal protein synthesis level via AS to decrease translation of a significant proportion of the transcriptome and produce the protein isoforms needed for adaptation to stresses. This strategy may allow plants to reduce their metabolic cost but also maintain a sufficient level of regulatory capacity via inclusion of alternative and disordered domains in stress-responsive proteins through AS. Although mechanistic details of such a process are not available in any organism at the moment, however, supporting evidence has just emerged from yeast. Two independent studies using yeast as a model have revealed that introns mediate fitness under stress conditions (nutrient starvation) by repressing ribosomal protein genes (for details see below) [44,45]. In addition, AS may not only diversify the regulatory capability of plant genes during initial stress episodes but also mediate crosstalk between a given metabolic state and protein diversity/abundance to cope with stressful conditions in the long term. Epigenetic modifications in plants such as DNA methylation and histone modifications define an epigenetic code that translates environmental stresses into an epigenetic footprint affecting cellular signalling network, and could also be recreated upon a recurring stress in the same or future generations [46]. In this way, AS may also be involved in stress memory mediated by epigenetic codes [47,48] and only after repeated onsets of similar stresses, plants could employ AS to generate more protein diversity or preserve the regulatory control in the long term [36,37].

Transcription and Splicing Dynamics in Plants

Transcription is a fundamental process to orchestrate gene expression patterns in response to different developmental and environmental cues. Surprisingly, limited information is

available on the mechanism of transcription in plants [49]. Human promotors are GC-rich [49,50], whereas plant promoters are AT-rich and tend to inhibit nucleosome formation, promoting DNA flexibility and transcription factor recruitment [51]. Comparison of RNA-seq and **global run-on sequencing (GRO-seq)** data sets in arabidopsis revealed a high correlation between nascent and steady-state transcripts [49]. Further, stable transcripts were associated with biological functions like translation, photosynthesis and metabolic functions. On the other hand, unstable transcripts had a higher representation of stimulus response genes, signal transduction, and hormones [49]. These results highlight that conserved genes associated with housekeeping functions are more stable compared with highly regulated transcripts. In view of these findings, it would be reasonable to speculate that AS transcripts, as a result of their dynamic nature, would be more suited for regulatory roles. Previous GRO-seq data showed that plant promoters lack promoter-proximal pausing of RNA polymerase II (RNAPII) and divergent transcription, which are prevalent in humans as well as yeast and drosophila [49,52]. However, very recent GRO-seq and **plant native elongating transcript sequencing (pNET-seq)** experiments from arabidopsis indicate that RNAPII with an unphosphorylated carboxyl-terminal domain (CTD) indeed accumulates downstream of transcription start sites (TSS) [53]. However, promoter-proximal pausing in arabidopsis is much more loose (broad peak) compared with mammals where pausing occurs in a narrow window of 25-50 nt [53]. These findings indicate that efficient RNAPII recruitment, as well as release from promoter-proximal pausing is necessary for efficient transcriptional response in arabidopsis. Interestingly, plant promoters also show **Ser2P CTD** RNAPII accumulation adjacent to the 3' polyadenylation site (PAS), suggesting the presence of a surveillance mechanism before transcription termination [53]. In vitro work in yeast proposed that RNAPII pausing after PAS may increase surveillance time and aid in mRNA degradation [54]. In addition, **Ser5P CTD** RNAPII elongates more slowly in exons compared with

introns to provide more time for the spliceosome to appropriately select splice sites in arabidopsis [53]. These data show that RNAPII CTD phosphorylation is a dynamic process and maybe even more important for sessile organisms like plants to maintain appropriate transcriptional and splicing dynamics under varied conditions. Since AS is largely co-transcriptional, distinctive features of plant transcription (transcription initiation and TSS/PAS proximal RNAPII pausing) may have a bearing on the transcriptional, splicing, and processing dynamics before a transcript is released from the transcription and splicing machinery [49,53,55].

Plant promoters have open chromatin structure compared with humans, however the relationship between DNA methylation and nucleosome occupancy in both species is very similar and may influence transcription and splicing processes mediated by RNAPII speed and splicing factors recruitment [56,57]. For example, recent evidence from arabidopsis shows that temperature-dependant AS correlates with changes at the chromatin level to regulate flowering time. Interestingly, Pajaro et al. have shown that H3 lysine 36 trimethylation (H3K36me3) strongly mark genes (96%) that undergo AS upon increasing the temperature from 16°C to 25°C [58]. Moreover, H3K36me3 was shown to play a crucial role in regulating AS of the flowering-time gene *FLOWERING LOCUS M* (*FLM*) upon elevated temperature [58,59]. In line with this data, the histone demethylase *JUMONJI C DOMAIN-CONTAINING PROTEIN 30* (*JMJ30*) and its homologue *JMJ32* remove the repressive H3K27me3 mark at the *FLOWERING LOCUS C* (*FLC*) to prevent precocious flowering at elevated temperatures [60]. Furthermore, dynamic chromatin landscapes under variable environmental and stress conditions have also been proposed to engender appropriate transcriptional and splicing responses in the short and long term [48,61]. Since plants continuously monitor their physiology and metabolism, we speculate

that cues from the environment during daily cycles of day and night, and RNAPII pausing near promoters and polyadenylation sites are important for appropriate transcriptional response and can also serve the role of a checkpoint that does not allow the release of newly synthesised transcripts before they are appropriately spliced, methylated and/or tagged for nuclear sequestration (transcripts with IR) and/or translation.

The correspondence between transcript and protein abundance should be taken into consideration because different levels of correlation between mRNAs and proteins were found in multiple organs and tissues in arabidopsis [62]. In addition, arabidopsis plants exposed to the microbe-associated molecular pattern elf18 showed poor correlation between transcription and translation patterns [63]. A recent comprehensive study in maize also revealed that about half of the highly abundant mRNAs are not represented at the protein level [64]. Intriguingly, syntenic and orthologus genes between maize and sorghum showed high expression level and were nine times more likely to produce proteins compared with nonsyntenic genes [64]. These findings indicate that highly expressed and conserved genes are more likely to be translated. However, composition of certain splice variants can also affect their translational potential, for example, IR in the 3' or 5' UTR can introduce *cis*-elements that influence stability or translation efficiency in humans [65]. Similary, it has been also demonstrated that plants use the 5'UTR as a sensor to to promote translation of some transcripts under stress conditions [66,67]. Therefore, any variation in the secondary structure of 5'UTRs via AS is likely to impact translational efficiency [66–68]. Furthermore, similarities in the *cis*-context and possibly the associated chromatin environment may also be important factors for mRNA and protein expression levels to achieve comparable gene expression and translation patterns between different species and/or growth or stress conditions. We posit that transcript variation may not be the sole controller of protein

diversity and abundance and plants may exercise a strong influence over these decisions taking into account their metabolic state, growth conditions, photosynthesis rates and status of sugar/starch reserves [69,70].

AS and Intrinsically Disordered Proteins/Regions: a way to regulate plants environmental fitness

Intrinsically disordered proteins or regions were termed as the junk proteome, however recent evidence shows they control important cellular functions via transcriptional regulation, cell cycle, chaperone formation and enrichment of regulatory capacity especially under stress conditions **(Figure 1)** [71]. Interestingly, highly conserved enzymes are normally not enriched in IDRs, whereas multifunctional enzymes contain disproportionately long IDRs [37]. Additionally, most eukaryotic proteins involved in transcription and RNA processing exhibit strong enrichment in IDRs that function in the formation of membraneless organelles in cells such as nuclear speckles, heterochromatin domains, stress granules and processing bodies [38,72,73]. Interestingly, stress granules can sequester and protect both RNAs and proteins from stress-induced damage [74,75] and alter signaling pathways during stress as shown for mammalian/mechanistic Target of Rapamycin Complex 1 (mTORC1) [76]. Recent data from two yeast studies demonstrate that introns are essential to promote resistance to stress conditions via the nutrient sensing TORC1 pathway [44,45]. In the first study [45], introns were found to be essential to downregulate ribosomal protein genes (RPGs) under starvation conditions to promote fitness in the wild type strains. Conversely, intron-deletion strains failed to survive under these conditions due to upregulation of RPGs and respiration-related genes, resulting in uncontrolled growth and starvation [45]. Intriguingly, excised introns, which are rapidly degraded under nutrient-rich conditions, accumulate as linear RNAs under stress conditions [44]. In the second study, deletion of these unusual

spliceosomal introns via the **CRISPR-Cas9 system** resulted in higher growth via TORC1 mediated stress response as well [44]. The presence of intron-mediated regulation of growth response in a eukaryote (yeast) is remarkable and it is tempting to speculate that similar mechanism exists in higher eukaryotes like plants, for at least, a subset of growth and stress-responsive genes.

Biased distribution of nucleotides at splice junctions is important for spliceosome recognition, however, most nucleotides at splice junctions and among *cis*-regulatory elements, code for disorder-promoting amino acids (Lysine, Glutamic acid and Arginine) [77]. Interestingly, exonic splicing enhancers are more prevalent in exons encoding disordered protein regions compared to exons associated with structured regions in many taxa including plants [77]. Since most protein segments affected by AS are often intrinsically disordered, these likely confer additional regulatory capacity by not only changing the three-dimensional structure but also their post-translational modifications (PTMs) to further diversify their function and substrate specificity in different cells under biotic and abiotic stress conditions in plants [36–39]. In general, the human proteome is more disordered, however genes involved in environmental responses are significantly more disordered in arabidopsis [78]. It is possible that the scheme of regulation via IDPs-AS-PTM is more relevant in plant species due to the prevalence of AS under stress conditions where a fine balance between photosynthesis, resource allocation, and acclimation response needs to be generated for adaptive responses and survival [37,79]. Under stress, plants display re-arrangement of their chromatin structure, which might also affect co-transcriptional splicing outcomes and differential splice site selection and increase AS diversity [80]. Recently, it has been shown that in addition to a regulatory role, IDPs play a central role in organisation and assembly of many macromolecular membraneless organelles including speckles, processing

bodies, stress granules and chromatin domains [35,73,78]. Consequently, IDPs might be a result of this stress-dependent chromatin modulation to help plants adapt in the short term. Stress- and stage-dependent IDPs can explain how the environment is capable of modulating the three-dimensional structure and PTMs of their proteins via AS. Hence, it is possible that IDPs provide condition-specific and enhanced regulatory network of transcriptional, splicing and translational regulators, and chaperones required for fine-tuning gene expression and refining the proteome in a given tissue under stressful conditions **(Figure 1)**. It has been proposed IDPs with AS and PTMs significantly contribute to the diversification of protein function and may also buffer against undesirable changes [37]. Furthermore, the presence of disordered regions in non-structural domains can aid neo-functionalization by evading the selection pressure that a protein with an altered structural domain would experience [36,37].

Translational Coincidence in Plants: The bright side of translation

Plants employ their internal, 24-hour timer, the "circadian clock", to synchronize daily activities to predictable changes in the environment [81], which provides a competitive advantage and maximizes productivity [82]. Evidence from previous studies shows that photosynthesis and starch synthesis rates during the day and resource mobilization to fuel growth during the night are tuned by the plant clock but are also dependent on the length of the photoperiod and growth in the previous night [83]. A prominent mechanism for clock control of physiological pathways is via the rhythmic regulation of RNA accumulation [81], including regulated AS [21,24,84,85]. Thousands of plant genes show rhythmic expression, with peaks across the day and night. These RNA rhythms (for mostly higher metabolic activity genes associated with photosynthesis, primary/secondary metabolism and pigment biosynthesis) interact with the photoperiod, where translation rate is higher during the light interval than in darkness [69,86]. Plants combine transcript rhythms and translational

247 regulation to tune protein expression in different photoperiods, *via* a mechanism called
248 **"Translational coincidence"**. For RNAs peaking late in the photoperiod, the higher
249 ribosome loading in the light interval only coincides with high mRNA levels during longer
250 photoperiods. If the photoperiod ends before the RNA level rises, daily protein synthesis
251 might, therefore, be lower. One way to increase levels of a protein under long photoperiods,
252 as in summer, is to time a rhythmic peak of RNA synthesis late in the day **(Figure 2)** [86].
253 arabidopsis proteome analysis in different photoperiods revealed that enzymes involved in
254 primary/secondary metabolism and photosynthesis were more abundant and plants show
255 higher metabolic activity under longer photoperiods [86]. Hundreds of proteins with rhythmic
256 RNAs peak late in the day were present at higher levels in these long photoperiod conditions,
257 whereas proteins with morning-peaking RNAs were more abundant in short photoperiods.

258 Since the timing of expression of a particular gene can influence its translation
259 patterns, it is logical to ask whether the same relationship holds true for alternatively spliced
260 transcripts. Indeed, light conditions regulate AS of SR30 pre-mRNA, which encodes a
261 serine/arginine-rich protein involved in RNA splicing in arabidopsis, and influence their
262 translation patterns [32]. One of the splice variants of SR30 (SR30.1) is rapidly generated
263 upon exposure to light and exported to the cytoplasm for translation as evident from the
264 abundance of SR30.1 protein [32]. In contrast, another splice variant, SR30.2 only appears in
265 dark-grown seedlings and is enriched in nuclear fractions with poor representation among
266 ribosome-associated transcripts. Interestingly, global analysis of AS in arabidopsis etiolated
267 seedlings exposed to different wavelengths of light revealed that many events switch from
268 probably unproductive variants in darkness to productive variants in light during seedling
269 photomorphogenesis [7]. Similarly, RS31 gene encoding another serine/arginine-rich splicing
270 factor in arabidopsis produces three isoforms under light conditions [87]. Of these, mRNA1

codes for the full-length protein and mRNA2 and mRNA3 are retained in the nucleus [87]. Interestingly, mRNA1 abundance considerably decreases under dark conditions without a significant drop in RS31 transcripts. Transgenic lines overexpressing mRNA1 show no phenotype under 16 and 8 hours of light and dark conditions, respectively, however result in yellowish and small seedlings under dark or low light intensity compared with WT or RS31 mutants as a result of lower levels of chlorophylls a and b [87]. Interestingly, plants treated with a drug that blocks electron transfer from **photosystem II** to the **plastoquinone** pool, mimics the effect of darkness on RS31 AS, indicating that a retrograde signal travels from the chloroplast to the nucleus. These data suggest that down-regulation of mRNA1 under dark conditions via AS is crucial for normal growth and development of arabidopsis plants under changing light conditions. Importantly, signals from chloroplast controlling nuclear events and a complex mechanism like AS is intriguing and indicates that environmental condition can influence gene regulatory mechanisms to confer plant fitness. However, it is notable that such crosstalk may take a long time to develop, considering the evolutionary history of chloroplasts and photosynthetic systems [88,89]. Alternative splicing of SR30 and RS31 genes can serve as a powerful model to understand why some splice variants appear only under variable environmental conditions and translated or retained in the nucleus. Additionally, these results support the notion that the metabolic state of a plant is closely regulated under different photoperiods and/or stress conditions, in part by altering which fraction of the transcriptome would be translated. Since AS transcripts are more abundant under stress condition, plants must tightly control what mRNA species will be translated to keep the metabolic cost of protein synthesis down [69,70]. It is therefore not surprising that a significant proportion of AS transcripts (IR) is either sequestered in the nucleus or degraded via the NMD pathway. Furthermore, since plants exhibit more protein translation under longer photoperiod (optimum energy supply) [86], we hypothesize that fewer proteins

(mostly IDPs) derived via AS under stress (limited energy supply) become a preferred choice to maintain essential regulatory control with minimum energy cost. Clearly, further work using ribosomal foot-printing and/or Mass Spec (see Box 1) techniques needs to be done to illuminate this phenomenon [90,91].

Concluding Remarks and Future Perspectives

All life forms need to orchestrate their transcriptome patterns to produce an appropriate response under normal and stress conditions. However, plant transcriptomes need to promote efficient carbon fixation and its utilization during the diurnal cycle at different growth and developmental stages. Therefore, it is intriguing that plants generate more splicing variation under stress conditions to fine-tune their gene expression patterns. It is therefore unlikely that plants would produce more proteins under limited energy supply [63,64]. Additionally, AS transcripts can produce nonsense transcripts and would result in truncated proteins if translated [42,92–94] **(Figure 1)**. Similarly, most IR transcripts, if translated, would produce proteins with IDRs and may not confer any specific function. However, most IR transcripts are trapped in the nucleus and thus remain untranslated [29]. Therefore, plants employ AS to not only alter their transcriptional response but also to influence proteome composition via sequestration of intron-containing RNAs and other alternatively spliced transcripts. It is also possible that similar to yeast [45], plant spliceosomal introns also play regulatory roles under stress conditions, however further work is needed to illuminate this phenomenon. Alternatively, plants may generate additional regulatory capacity via translating some of the AS transcripts that harbour IDRs in different transcription factors including clock genes, and splicing factors to confer enhanced regulatory capacity to interact with multiple partners, enzymes and their substrates [36,37,71,78]. This is reminiscent of *Down syndrome cell adhesion molecule* (*Dscam*) protein, which is required for neuronal connections in drosophila

[95]. *Dscam* gene can generate thousands of splice isoforms. Although, all splice isoforms share the same domain, variable amino acids with in the immunoglobulin (Ig) domains confer binding specificity and contribute to complex neuronal wiring [95,96]. In this way, isoform diversity provides each neuron with a unique identity to facilitate self-recognition, which is essential for neuronal wiring in drosophila [95,96].

We also propose that AS increases regulatory capacity in the short term but only contributes to protein diversity in the long term when different combinations have been tried over many generations and purifying selection has taken its course [37,77,97]. A recent study showed that plants possess splicing memory for heat stress and only previously primed plants with heat stress show a predicted AS response to the same stress again [47]. This short-term AS memory may be engendered through specific chromatin marks that in turn give birth to long-term adaptations mediated by chromatin landscape. This strategy provides spatiotemporal order and reproduction of a specific AS pattern under a similar condition, tissue and/or developmental stage [48]. Since chromatin state also mediates transcription and splicing dynamics [80,98,99], chromatin environment may not only mediate specific AS outcomes but could also serve as an epigenetic footprint to trigger a comparable response in the event of a similar stress in the future [47,48,99]. We envisage that understanding the transcriptional and translational dynamics of different AS transcripts in concert with associated chromatin marks, in different photoperiods and environmental conditions will be fruitful to understand the impact of AS on the alternative proteome. To fully appreciate the role of AS in gene regulation and protein diversity, we need to not only understand the chromatin context in which different AS patterns appear in the short and long term but also look at their partners by using yeast hybrid system and modified MS and **LC-MS** techniques

in a tissue and condition-specific manner among diverse populations and under different conditions (see also outstanding questions).

Acknowledgements

This work was supported by funding from The Leverhulme Trust (RPG-2016-014). We also thank Andreas Wachter and three anonymous reviewers for their helpful comments and constructive criticism.

REFERENCES

1 Zhu, J.-K. (2016) Abiotic stress signaling and responses in plants. *Cell* 167, 313–324

2 Filichkin, S.A. *et al.* (2015) Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant* 8, 207–227

3 Ding, F. *et al.* (2014) Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genomics* 15, 1–14

4 Kwon, Y.-J. *et al.* (2014) Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC Plant Biol.* 14, 136

5 Yu, H. *et al.* (2016) Transcriptome survey of the contribution of alternative splicing to proteome diversity in Arabidopsis thaliana. *Molecular Plant* 9, 749–752

6 Marquez, Y. *et al.* (2015) Unmasking alternative splicing inside protein-coding exons

defines exitrons and their role in proteome plasticity. *Genome Res.* 25, 995–1007

7 Hartmann, L. *et al.* (2016) Alternative splicing substantially diversifies the transcriptome during early photomorphogenesis and correlates with the energy availability in Arabidopsis. *Plant Cell* 28, 2715-2734

8 Reddy, A.S.N. *et al.* (2013) Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683

9 Sterne-Weiler, T. *et al.* (2013) Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* 23, 1615–1623

10 Floor, S.N. and Doudna, J.A. (2016) Tunable protein synthesis by transcript isoforms in human cells. *Elife* 5,

11 Weatheritt, R.J. *et al.* (2016) The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123

12 Yang, X. *et al.* (2016) Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* 164, 805–817

13 Liu, Y. *et al.* (2017) Impact of alternative splicing on the human proteome. *Cell Rep.* 20, 1229–1241

14 Kahles, A. *et al.* (2018) Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 34, 211–22

15 Brosch, M. *et al.* (2011) Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res.* 21, 756–767

16 Tress, M.L. *et al.* (2007) The implications of alternative splicing in the ENCODE protein complement. *Proc. Natl. Acad. Sci.* 104, 5495–5500

17 Tress, M.L. *et al.* (2008) Proteomics studies confirm the presence of alternative protein isoforms on a large scale. *Genome Biol.* 9, R162

18 Abascal, F. *et al.* (2015) Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11, e1004325

19 Tress, M.L. *et al.* (2017) Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* 42, 98–110

20 Gibon, Y. *et al.* (2009) Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. *Plant Cell Environ.* 32, 859–874

21 Filichkin, S. *et al.* (2015) Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Current Opinion in Plant Biology*, 24, 125–135

22 Filichkin, S.A. *et al.* (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res.* 20, 45–58

23 Marquez, Y. *et al.* (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22, 1184–1195

24 Filichkin, S.A. and Mockler, T.C. (2012) Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes. *Biol. Direct* 7, 20

25 Lejeune, F. *et al.* (2004) eIF4G is required for the pioneer round of translation in mammalian cells. *Nat. Struct. Mol. Biol.* 11, 992–1000

26 Maquat, L.E. *et al.* (2010) The pioneer round of translation: features and functions. *Cell* 142, 368–374

27 Durand, S. and Lykke-Andersen, J. (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat. Struct. Mol. Biol.* 20, 702-709

28 Rufener, S.C. and Mühlemann, O. (2013) EIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* 20, 710-717

29 Gohring, J. *et al.* (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

30 Sun, S. *et al.* (2010) SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat. Struct. Mol. Biol.* 17, 306–312

31 Wei, G. *et al.* (2018) Position-specific intron retention is mediated by the histone methyltransferase SDG725. *BMC Biol.* 16, 44

32 Hartmann, L. *et al.* (2018) Subcellular compartmentation of alternatively-spliced transcripts defines SERINE/ARGININE-RICH PROTEIN 30 expression. *Plant Physiol.* 176, 2886-2903

33 Staiger, D. and Simpson, G.G. (2015) Enter exitrons. *Genome Biol.* 16, 136-138

34 Johnson, D.S. *et al.* (2007) Genome-wide mapping of in vivo protein-DNA

interactions. *Science* 316, 1497–1502

35 Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83, 553–584

36 Niklas, K.J. *et al.* (2015) Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell Dev. Biol.* 3, 1–13

37 Niklas, K.J. *et al.* (2018) The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* 69, 1437–1446

38 Strom, A.R. *et al.* (2017) Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245

39 Buljan, M. et al. (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46, 871–883

40 Drechsel, G. et al. (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

41 Brogna, S. *et al.* (2016) The Meaning of NMD: translate or perish. *Trends Genet.* 32, 395–407

42 Trcek, T. *et al.* (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev.* 27, 541–51

43 Shaul, O. (2015) Unique aspects of plant nonsense-mediated mRNA decay. *Trends Plant Sci.* 20, 767–779

447 44 Morgan, J.T. *et al.* (2019) Excised linear introns regulate growth in yeast. *Nature* DOI:
448 10.1038/s41586-018-0828-1

449 45 Parenteau, J. *et al.* (2019) Introns are mediators of cell response to starvation. *Nature*
450 DOI: 10.1038/s41586-018-0859-7

451 46 Lang-Mladek, C. *et al.* (2010) Transgenerational inheritance and resetting of stress-
452 induced loss of epigenetic gene silencing in arabidopsis. *Mol. Plant* 3, 594–602

453 47 Ling, Y. *et al.* (2018) Thermopriming triggers splicing memory in Arabidopsis. *J. Exp.*
454 *Bot.* 69, 2659–2675

455 48 Lämke, J. and Bäurle, I. (2017) Epigenetic and chromatin-based mechanisms in
456 environmental stress adaptation and stress memory in plants. *Genome Biol.* 18, 124

457 49 Hetzel, J. *et al.* (2016) Nascent RNA sequencing reveals distinct features in plant
458 transcription. *Proc. Natl. Acad. Sci.* 113, 12316–12321

459 50 Core, L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and
460 divergent initiation at human promoters. *Science* 322, 1845–1848

461 51 Zuo, Y.C. and Li, Q.Z. (2011) Identification of TATA and TATA-less promoters in
462 plant genomes by integrating diversity measure, GC-Skew and DNA geometric
463 flexibility. *Genomics* 97, 112–120

464 52 Nechaev, S. *et al.* (2010) Global analysis of short RNAs reveals widespread promoter-
465 proximal stalling and arrest of Pol II in Drosophila. *Science* 327, 335–338

466 53 Zhu, J. *et al.* (2018) RNA polymerase II activity revealed by GRO-seq and pNET-seq
467 in Arabidopsis. *Nature Plants* 12, 1112-1123

468 54 Anamika, K. *et al.* (2012) RNA polymerase II pausing downstream of core histone
469 genes is different from genes producing polyadenylated transcripts. *PloS one* 7,
470 e38769

471 55 Irimia, M. *et al.* (2014) A highly conserved program of neuronal microexons is
472 misregulated in autistic brains. *Cell* 159, 1511–1523

473 56 Chodavarapu, R.K. *et al.* (2010) Relationship between nucleosome positioning and.
474 *Nature* 466, 1–5

475 57 Naftelberg, S. *et al.* (2015) Regulation of alternative splicing through coupling with
476 transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198

477 58 Pajoro, A. *et al.* (2017) Histone H3 lysine 36 methylation affects temperature-induced
478 alternative splicing and flowering in plants. *Genome Biol.* 18, 102

479 59 Steffen, A. and Staiger, D. (2017) Chromatin marks and ambient temperature-
480 dependent flowering strike up a novel liaison. *Genome Biol.* 18, 119

481 60 Gan, E.S. *et al.* (2014) Jumonji demethylases moderate precocious flowering at
482 elevated temperature via regulation of FLC in Arabidopsis. *Nat. Commun.* 5, 5098

483 61 Brzezinka, K. *et al.* (2016) Arabidopsis FORGETTER1 mediates stress-induced
484 chromatin memory through nucleosome remodeling. *Elife* 5, e17061

485 62 Baerenfaller, K. *et al.* (2008) Genome-scale proteomics reveals Arabidopsis thaliana
486 gene models and proteome dynamics. *Science* 320, 938-941

487 63 Xu, G. *et al.* (2017) Global translational reprogramming is a fundamental layer of
488 immune regulation in plants. *Nature* 545, 487-490

64 Walley, J.W. *et al.* (2016) Integration of omic networks in a developmental atlas of maize. *Science* 353, 814-818

65 Jacob, A.G. and Smith, C.W.J. (2017) Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 136, 1043–1057

66 Remy, E. *et al.* (2014) Intron retention in the 5′UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in Arabidopsis. *PLoS Genet.* 10, 15–19

67 Álvarez, D. *et al.* (2016) Carotenogenesis is regulated by 5′ UTR-mediated translation of phytoene synthase splice variants. *Plant Physiol.* 172, 2314–2326

68 Matsuura, H. *et al.* (2013) A computational and experimental approach reveals that the 5′-proximal region of the 5′-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in arabidopsis. *Plant Cell Physiol* 54, 474-483

69 Piques, M. *et al.* (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Mol. Syst. Biol.* 5, 314

70 Ishihara, H. *et al.* (2017) Growth rate correlates negatively with protein turnover in Arabidopsis accessions. *Plant J.* 91, 416–429

71 Dunker, A.K. *et al.* (2013) What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disord. Proteins* 1, e24157

72 Minezaki, Y. *et al.* (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359, 1137–1149

73 Rai, A.K. *et al.* (2018) Kinase-controlled phase transition of membraneless organelles

in mitosis. *Nature* 559, 211–216

74 Chavali, S. *et al.* (2017) Intrinsically disordered proteins adaptively reorganize cellular matter during stress. *Trends Biochem. Sci.* 42, 410-412

75 Riback, J.A. *et al.* (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* 168, 1028-1040

76 Wippich, F. *et al.* (2013) Dual specificity kinase DYRK3 couples stress granule condensation/ dissolution to mTORC1 signaling. *Cell* 152, 791–805

77 Smithers, B. *et al.* (2015) Splice junctions are constrained by protein disorder. *Nucleic Acids Res.* 43, 4814–4822

78 Pietrosemoli, N. *et al.* (2013) Genome-wide analysis of protein disorder in Arabidopsis thaliana: implications for plant environmental adaptation. *PLoS One* 8, e55524

79 Bah, A. and Forman-Kay, J.D. (2016) Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* 291, 6696–705

80 Ullah, F. *et al.* (2018) Exploring the relationship between intron retention and chromatin accessibility in plants. *BMC Genomics* 19, 21

81 Millar, A.J. (2016) The intracellular dynamics of circadian clocks reach for the light of ecology and evolution. *Annu. Rev. Plant Biol.* 67, 595–618

82 Seo, P.J. and Mas, P. (2015) STRESSing the role of the plant circadian clock. *Trends Plant Sci.* 20, 230–237

83 Graf, A. *et al.* (2010) Circadian control of carbohydrate availability for growth in

Arabidopsis plants at night. *Proc. Natl. Acad. Sci.* 107, 9458–9463

84 James, A.B. *et al.* (2012) Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell* 24, 961–81

85 Schmal, C. *et al.* (2013) A circadian clock-regulated toggle switch explains AtGRP7 and AtGRP8 oscillations in Arabidopsis thaliana. *PLoS Comput. Biol.* 9, e1002986

86 Seaton, D.D. *et al.* (2018) Photoperiodic control of the *Arabidopsis* proteome reveals a translational coincidence mechanism. *Mol. Syst. Biol.* 14, e7962

87 Petrillo, E. *et al.* (2014) A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* 344, 427-430

88 Baena-González, E. *et al.* (2007) A central integrator of transcription networks in plant stress and energy signalling. *Nature* 448, 938-942

89 Xiong, J. and Bauer, C.E. (2002) Complex evolution of photosynthesis. *Annu. Rev. Plant Biol* 53, 503-521

90 Mustroph, A. *et al.* (2009) Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *Proc. Natl. Acad. Sci.* 106, 18843-18848

91 Juntawong, P. *et al.* (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc. Natl. Acad. Sci.* 111, E203-E212

92 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

551 93 Kalyna, M. *et al.* (2012) Alternative splicing and nonsense-mediated decay modulate
552 expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40, 2454-
553 2469

554 94 Sato, H. *et al.* (2008) Efficiency of the pioneer round of translation affects the cellular
555 site of nonsense-mediated mRNA decay. *Mol. Cell* 29, 255–262

556 95 Wojtowicz, W.M. *et al.* (2004) Alternative splicing of Drosophila Dscam generates
557 axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118,
558 619-633

559 96 Hattori, D. *et al.* (2007) Dscam diversity is essential for neuronal wiring and self-
560 recognition. *Nature* 449, 223

561 97 Kovacs, E. *et al.* (2010) Dual coding in alternative reading frames correlates with
562 intrinsic protein disorder. *Proc. Natl. Acad. Sci.* 107, 5429–5434

563 98 Luco, R.F. *et al.* (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26

564 99 Liu, H. *et al.* (2018) Distinct heat shock factors and chromatin modifications mediate
565 the organ-autonomous transcriptional memory of heat stress. *Plant J.* 95, 401-413

566 100 Olsen, J. V. *et al.* (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine
567 residues. *Mol. Cell. Proteomics* 3, 608–614

568 101 Wang, X. *et al.* (2018) Detection of proteome diversity resulted from alternative
569 splicing is limited by trypsin cleavage specificity. *Mol. Cell. Proteomics* 17, 422–430

570 102 Ning, K. and Nesvizhskii, A.I. (2010) The utility of mass spectrometry-based
571 proteomic data for validation of novel alternative splice forms reconstructed from

RNA-Seq data: A preliminary assessment. *BMC Bioinformatics* 11, S14

103 Sheynkman, G.M. *et al.* (2013) Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. *Mol. Cell. Proteomics* 12, 2341–53

104 Wang, X. *et al.* (2013) CustomProDB: An R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* 29, 3235–3237

105 Ingolia, N.T. (2016) Ribosome footprint profiling of translation throughout the genome. *Cell* 165, 22–33

106 Inada, T. (2017) The ribosome as a platform for mRNA and nascent polypeptide quality control. *Trends Biochem. Sci*. 42, 5–15

Box1: Limitations to Detect Alternative Isoforms at the Proteome Level

In the shotgun proteomic analysis, proteins are first digested proteolytically into smaller peptides using trypsin and subsequently analysed by LC-MS/MS [100]. Trypsin, the most common enzyme used in Mass Spec cleaves at the C-terminus of lysine or arginine to produce peptides with optimal length and charge [100]. Peptides spanning exon-exon junctions provide direct evidence of splice variants at the protein level. Interestingly, lysine and arginine are the most enriched amino acids at exon-ending or exon-exon junctions of transcripts [101]. Exon-exon junctions are preferred sites for trypsin digestion, hindering detection of junction-specific peptides and identification of novel alternative splicing peptides

in the proteo-genomics analysis [102–104]. To overcome trypsin digestion limitations, specificity of five proteases including Lys-C, Glu-C, chymotrypsin, Asp-N, and Arg-C was evaluated recently [101]. Among these five enzymes, the highest number of detectable junctions including exon-ending and exon-exon junctions were observed in chymotrypsin digestion, making it a protease of choice in LC-MS/MS studies, especially to predict RNA splicing derived peptides [101]. Since different protein isoforms of the same gene may be localized in different tissues conferring diverse physiological outcomes, it would be useful to improve the sensitivity of current proteomic analysis methods. Alternatively, ribosome profiling/foot-printing along with next-generation sequencing (NGS) (Ribo-Seq), can be employed as an alternative strategy to use ribosome bound transcripts as a proxy for translation [91,105]. However, foot-printing data should be treated with caution as ribosome bound transcripts may not be translated as a result of ribosomal scrutiny during the pioneer round of translation [106]. In the future, quantitative Ribo-Seq and proteomic data from multiple tissues in the context of RNA-metabolism, degradation, and other features may help to improve the efficiency to detect translated transcripts.

Legends

Figure 1: A hypothetical schematic diagram showing fates of alternatively spliced transcripts under normal and stress conditions in plants. AS generates multiple transcripts under normal (N1-N4) as well as stress (S1-S5) conditions. Constitutively spliced transcripts (N1 and S1) and alternatively spliced PTC- transcripts (N3 and S2) are translated into functional protein isoforms (FP_s) and intrinsically disordered proteins (IDP_s). Alternatively spliced PTC+ transcripts (N2, N4, S3, and S4) are either degraded via the NMD pathway (N4 and S4) or escape NMD (S3) to generate truncated proteins (TP_s). Although present in both conditions, FP_s are more abundant under normal conditions, whereas TP_s and IDP_S constitute the majority of stress-induced proteome.

Figure 2. Translational coincidence upon photoperiod length and long-term changes. Under long photoperiods (day-time represented by yellow colour), plants translate a higher proportion of their transcriptome to produce more proteins, to support a higher degree of metabolic activity. However, under a short photoperiod (evening and night-time represented by light and dark blue colour, respectively), ribosome loading and translational efficiency are reduced as a result of lower demand. In this way, plants may modulate their proteome using the same transcriptomic pool upon varied physiological needs. Moreover, during different growth stages (A-B-C), the relationship between transcript abundance and protein diversity may not be linear to maintain desirable cost to benefit ratio and regulatory capacity.

GLOSSARY

Alternative Splicing: A gene regulatory mechanism that produces different messenger-RNAs (mRNAs) from a single gene via inclusion and/or exclusion of exons or introns fully or partially in different transcripts.

Mass-spectrometry (MS): An analytical technique to identify small molecules and macromolecules (including proteins) on the basis of mass to charge ratio.

Liquid chromatography-MS (LC-MS): A technique that combines the power of liquid chromatography for sample ionization/physical separation with MS.

Intron Retention: An alternative splicing event that retains an intron in the transcript.

Intrinsically Disordered Proteins: Proteins that lack well-defined globular three-dimensional structures and frequently interact with or function as hubs in protein interaction networks.

Intrinsically Disordered Region: Some proteins completely disordered, whereas others only harbour disordered sequences, referred to as intrinsically disordered regions (IDRs).

Translational Coincidence: Differences in the rates of protein synthesis across photoperiods that explain the changes in the coincidence of rhythmic RNA expression with light resulting in higher rates of translation.

Photosystem II: First protein complex located in the thylakoid membrane of chloroplasts that uses energy from sunlight to extract electrons from water molecules.

Plastoquinone: Carriers of electrons in Photosystem II that establish the electron transport chain during photosynthesis.

GRO-seq: Global run-on sequencing is a technique in which actively transcribing nascent RNAs are sequenced using next-generation sequencing platforms.

pNET-seq: Plant native elongating transcript sequencing is a technique that involves isolation of the 3' ends of actively transcribing genes via immunoprecipitation of the RNA polymerase II complex, to precisely map RNAPII position and is followed by next-generation sequencing.

Ser2(5)P CTD: The C-terminal domain (CTD) of the RNA polymerase II is dynamically phosphorylated during transcription via different phosphorylation patterns that help recruit required mRNA processing and histone modifying factors. Serines 2 (Ser2) and Ser5 are major phosphorylation sites in the CTD domain.

CRISPR-Cas9 system: CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9) is a naturally occurring bacterial derived genome editing system. CRISPR-Cas9 system allows insertion and deletion of genomic regions with greater precision than previously available methods.

OUTSTANDING QUESTIONS

1. To which extent alternatively spliced transcripts are engaged with the ribosomal machinery (partly known) and translated into proteins?

2. How do plants couple their AS events to photoperiodic changes to modulate their proteome upon physiological need through IDPs?

3. What is the impact of chromatin state on transcriptional dynamics, alternative splicing, epitranscriptome and translational efficiency of transcripts in plants?

4. To which extent PTC+ transcripts make truncated but functional proteins?

5. Similar to yeast, is there any regulatory role of plant spliceosomal introns under stress conditions?

Fig 1

Click here to access/download

Original Figure File

Figure 1.pptx

Fig 2

Click here to access/download

Original Figure File

Figure 2.pptx

Figure 1