Title: Impact of alternative splicing on Arabidopsis proteome

Andres V Reyes[1,2], Ruben Shrestha[1], TaraBryn S Grismer[1,2], Danbi Byun[1], Shou-Ling Xu[1,2]

[1]Division of biosphere science and engineering, Carnegie Institution for Science, Stanford, California, USA
[2]Carnegie Mass Spectrometry Facility, Carnegie Institution for Science, Stanford, California, USA

Alternative splicing is an important regulatory process in eukaryotes. In plants, the major form of alternative splicing is intron retention. Despite its importance, the global impact of AS on the Arabidopsis proteome has not been investigated. In this study, we address this gap by performing a comprehensive integrated analysis of how changes in AS can affect the Arabidopsis proteome using mutants that disrupt ACINUS and PININ, two evolutionarily conserved alternative splicing factors. We used tandem mass tagging (TMT) with real-time search MS3 (RTS-SPS-MS3) coupled with extensive sample fractionations to achieve very high coverage and accurate protein quantification. We then integrated our proteomic data with transcriptomic data to assess how transcript changes and increased intron retention (IIR) affect the proteome. For differentially expressed transcripts, we have observed a weak to moderate correlation between transcript changes and protein changes. Our studies revealed that some IIRs have no effect on either transcript or protein levels, while some IIRs can significantly affect protein levels. Surprisingly, we found that IIRs have a much smaller effect on increasing protein diversity. Notably, the increased intron retention events detected in the double mutant are also detected in the WT under various biotic or abiotic stresses. We further investigated the characteristics of the retained introns. Our extensive proteomic data help to guide the phenotypic analysis and reveal that collective protein changes contribute to the observed phenotypes of the increased anthocyanin, pale green, reduced growth, and short root observed in the *acinus pnn* double mutant. Overall, our study provides insight into the intricate regulatory mechanism of intron retention and its impact on protein abundance in plants.

Introduction

Next-generation sequencing (NGS)-based expression analyses have shown that more than 95% and 70% of multi-exon genes in animals and plants, respectively, undergo alternative splicing (AS) (1-3). Alternative splicing (AS) has been shown to be an important mechanism in gene regulation and is predicted to increase protein diversification.

Intron retention, a form of alternative splicing in which an intron remains in the mature transcript, has been studied in both plants and animals. In animals, although it is not the most dominant AS form, intron retention has been shown to be a relevant layer in the regulation of gene expression (4-6). In plants, intron retention is the most dominant form of AS in plants, accounting for more than 60% of AS. Intron retention has been associated with several response pathways, including developmental stages, responses to abiotic and biotic stresses, and the regulation of circadian rhythms, or dark to light transitions (3,7,8)

Despite its importance, the impact of AS, especially intron retention, at the proteome level remains elusive. In animals, most retention events occur in non-coding regions of transcripts, whereas in plants, intron retention occurs predominantly within coding regions, presumably affecting protein diversity. Interestingly, a substantial fraction of retained introns in plants are characterized as post-transcriptionally spliced (pts) introns. The transcripts containing pts introns are characterized by chromatin tethering and polyadenylation, similar to their animal counterparts known as detained introns (8,9). It is hypothesized that upon receiving environmental cues, these introns are spliced out and their final transcripts are exported for translation, adding another layer to the regulatory dynamics affecting the proteome (8). However, since the half-life of retained introns is expected to be over 1 hour (4), the impact of IR remains to be determined. In addition, the intron-retained transcript may be transported to the cytosol, translated, but trigger nonsense-mediated decay. Alternatively, if these transcripts escape NMD, truncated proteins may be produced, potentially further increasing protein diversity.

In this study, we aim to fill a critical gap in understanding by performing a comprehensive, integrated analysis of how alterations in alternative splicing (AS) can affect the Arabidopsis proteome and use proteome changes to predict phenotype. Specifically, we study the effects of mutations that disrupt ACINUS and PININ, two highly conserved AS factors that affect hundreds of gene expressions and alternative splicing. To achieve this, we use tandem mass tagging (TMT)

with real-time search MS3 (RTS-SPS-MS3) in conjunction with extensive sample fractionations, which allows us to achieve extensive coverage and accurate protein quantification. We then integrate our proteomic findings with transcriptomic data to evaluate the effects of transcript changes and increased intron retention (IIR) on the proteome. Overall, our comprehensive study indicates an intricate regulatory mechanism of intron retention and shows a complex impact on protein abundance in plants.

Results

Intron retention events detected in *acinus pinin* mutants are physiologically regulated in WT upon abiotic or biotic stress

We have previously shown that ACINUS and PININ play a role in the regulation of both transcription and alternative splicing. Using RackJ, we detected an increase of 258 intron retention events in 225 genes in the *acinus pinin* double mutant (10). To investigate whether these intron retentions could be physiologically replicated in WT, we data-mined the plant intron retention database (plantintron.com), which contains over 25,000 Arabidopsis RNA-seq libraries (7), prioritizing similar experiments performed by independent groups or in different projects. We found that several intron retention events in the *acinus pinin* double mutant are similarly retained in WT under biotic and abiotic stress. For example, the retention of intron 16 of SGR6 is cold induced (Fig. 1A); similarly, intron 9 of AT1G34340 is also significantly retained by cold treatments and by osmotic stress (Fig. 1B). Retention of intron 6 of AT1G65540 and intron 16 of AT3G59660 is significantly induced by flagellin (flg22) treatment (Fig. 1C and Fig. 1D), whereas intron 4 of AT3G27930 is significantly induced by either dehydration or heat (Fig.1E). This suggests that the intron retention detected in the *acinus pinin* double mutant may occur in the WT in response to environmental cues, highlighting that intron retention events are physiologically regulated in plants.

Multi omics integrated workflow to study transcription, alternative splicing and proteome

Since both transcription and alternative splicing are perturbed in the *acinus pinin* mutant, we developed an integrated pipeline to answer several questions: 1) Do the altered transcript levels affect protein abundance? 2) To what extent are AS transcripts translated and have functional consequences at the proteome level in plants? Does IR contribute to protein abundance or even to proteome diversity? 3) Can we predict phenotype from changes in protein abundance? To systematically address these questions, we acquired new proteomics data and integrated our previous RNA-seq transcriptomics data (10). To achieve comprehensive coverage and accurate

protein quantification, we used the tandem mass tag (TMT) method (11). This approach allows for multiplexed sample measurements and extensive fractionation, which is optimized for high coverage and high quantification accuracy (Fig.2A). To ensure the robustness of the data, we included five biological replicates for both wild type (WT) and *acinus pinin* (*ap*) double mutants (Fig 2A).

Proteins extracted from WT (5 replicates) and *acinus pinin* mutants (5 replicates) were digested into peptides using trypsin and labeled using the TMT11 plex kit (Fig 2A), with the last channel of the mixed sample. The combined samples were subjected to reverse HPLC fractionation and LC-MSMS analysis. To improve the accuracy of quantification, we used a narrow isolation window of 0.7 Da for the precursor and implemented a real-time library search method (RTS-SPS-MS3) (12). This method improves sensitivity and accuracy of quantification by allowing only library-matched TMT-labeled peptides for MS3 acquisition (RTS) and by allowing the synchronous isolation of TMT-labeled MS2 fragment ions for MS3 fragmentation (synchronous precursor selection (SPS)-MS3) during real-time data acquisition.

High coverage achieved with TMT quantitative analysis

Using TMT quantitative analysis approaches, we achieved remarkably high coverage, quantifying a total of 13,850 protein groups and 108,350 peptides (Supplementary Table 1), with a false discovery rate (FDR) of 1% at both the protein and peptide levels (Fig 2B). Principal component analysis (PCA) further underscored the robustness of our TMT pipeline, with a clear distinction between WT and mutant groups, separated from the mixed channel (Fig. 2C). For clarity of quantification, we excluded 829 protein groups with peptides matching multiple genes. After consolidating groups matching different isoforms (1460 groups removed), we retained those with the highest peptide counts. The final number of protein groups quantified for unique genes was 11,561 (Supplementary Table 1). Our analysis identified 431 up-regulated proteins and 380 down-regulated proteins in the *acinus pinin* mutant (Fig 2B, 2D) with a 2-fold cut-off and an adjusted p-value of 0.05. Notably, the down regulated proteins include known targets, such as ACINUS, PININ, SR45 and SAP18 (10).

Low to modest correction between transcriptomic and proteomic changes.

To investigate whether changes in gene expression are proportionally reflected in changes in protein abundance, we compared the transcript changes to protein changes. Using RNA-seq, we previously showed that the *acinus pinin* double mutants have significantly decreased expression levels for 786 genes and increased expression levels for 767 genes (fold change >=2) compared to WT out of the 22,218 genes we quantified (10). We divided genes into three groups, upregulated, downregulated, and unchanged, using a twofold cutoff, and examined their changes in protein abundance using Pearson's correlation test.

Among the 767 genes with increased gene expression in *acinus pnn* mutants (≥ 2-fold), 247 genes were quantified at the proteome level (Fig 3A). Among them, 161 proteins (65.1%) showed a significant increase in protein levels (≥ 2- fold difference, p value = 0.05), while 85 proteins (34.4%) showed no significant change; only one protein contradicted the transcriptomic shift (Fig. 3A). To analyze the correlation between transcript and protein changes, we found a moderate correlation with a Pearson correlation coefficient (n= 247, R=0.39, p=1.6e-10, Pearson test) (Fig. 3B). Gene ontology (GO) analysis of the 161 proteins revealed several significantly enriched processes, including systemic acquired resistance (SAR), biosynthesis of anthocyanin-containing compounds, and response to bacteria (Fig. 3C, Supplementary Table 2). We observed that for some genes, the fold difference between WT and *acinus pinin* mutant is much higher at the protein level than at the transcript level, such as pathogenesis-related proteins (PR2, PR4, and PR5).

Of the 786 genes with decreased expression in *acinus pnn* mutant (≤0.5 fold), 234 genes were quantified at the proteome level (Fig. 3D). Of these, 130 proteins (55.6%) showed a significant decrease in protein levels in *acinus pinin* mutant (fold difference ≤ 0.5, p value = 0.05), while 101 proteins (43.2 %) showed no significant change. Three proteins contradicted the transcriptomic shift, appearing increased protein level despite decreased transcript levels (Fig. 3D and 3E). When we analyzed the correlation between transcript changes and protein changes for these genes, we found a weak Pearson correlation coefficient (n= 234, R=0.35, p value 2.9e-08, Pearson test) (Fig. 3E). GO analysis of these down regulated proteins (Fig 3F, Supplementary Table 2) shows that several processes are significantly enriched, including the suberin biosynthetic process, which is consistent with the previous correlation of *acinus pinin* double mutant with ABA hypersensitivity (10,13,14).

In addition, we also examined proteins corresponding to genes with no observed changes in RNA levels. Of the 20,662 quantified genes showing no change in the RNA-seq data, 10,962 were

quantified at the proteome level (Fig. 3G), of which 10,474 (95.5%) showed no significant proteome changes. The Pearson correlation coefficient is modest (n= 10,959, R=0.5, p-value= 2.2e-16, see Methods) (Fig. 3H). This indicates that most genes without changes at the transcript level do not show changes at the protein level.

However, there are some exceptions: 241 proteins (2.2%) show a significant increase and 247 proteins show a significant decrease (2.3%). For example, a leucine-rich receptor kinase (AT5G59670, LRRK) and CALMODULIN LIKE 10 (CML10) proteins show a significant increase in protein level, but no difference in transcript level (Fig 3H). Other proteins, such as SHOOT GRAVITROPISM 6 (SGR6) and GDA/CD39 nucleoside phosphatase (AT1G14250), SR45 and SAP18, show greatly reduced protein levels, although the RNA level is not affected. The dramatic reduction of SR45 and SAP18 is consistent with our previous SILIA/PRM targeted quantification (10), showing that the stability of SR45 and SAP18 is dependent on ACINUS and PININ. GO analysis of up-regulated proteins shows enrichment in defense-related processes (Fig. 3I, Supplementary Table 2), but the same GO analysis of down-regulated proteins shows no obvious enrichment (Supplementary Table 2).

Intron retention reduces protein levels

To understand how intron retention contributes to the transcriptome and proteome, we next focused on those genes and the proteins they encode that showed increased intron retention in the *acinus pinin* mutant. Previously (10), using RackJ, we identified an increase of 258 introns in 225 genes in the *acinus pinin* double mutant, of which 236 (91.5%) of these introns were enriched in the protein-coding region, including 231 that introduce a premature stop codon (PTC) and 5 that are exitrons. 23 are not in the protein coding region, 20 in the 5'UTR and 2 in the 3'UTR. This observed pattern contrasts with IRs detected in human and mouse tissues, where they tend to be enriched in untranslated regions (UTRs) and non-coding RNAs, while depleted in protein coding regions (5,15).

Since the vast majority of IR events fall within the coding sequence, one might wonder what their effect on the proteome is. Of these 225 intron-retained genes, we were able to quantify 188 proteins encoded by them (Fig. 4A). It is currently thought that some intron-retained scripts are subject to post-transcriptional splicing (PTS) regulation. They are tethered and retained in the nucleus, which may or may not reduce RNA in the cytosol, and thus affect the protein levels in unpredictable ways. Alternatively, transcripts with intron retention are transported to the cytosol

and translated, but induce nonsense-mediated decay (NMD), resulting in reduced protein abundance. Alternatively, the transcripts may escape the NMD, but the premature stop codon would result in a truncated protein (6).

To distinguish between these different outcomes, we analyzed each protein at the peptide quantification level. Briefly, for each protein, we separated the peptides before and after the intron retention site, referred to as N-terminal peptides (colored in red) and C-terminal peptides (colored in blue) relative to the predicted intron retention site (Fig. 4B). We reasoned that if the intron retention causes the overall difference in protein abundance, then both the N-terminal and the C-terminal peptides will show reduction. If the intron retention causes truncation of proteins, then the N-terminal peptides may show no change or a reduction, but the C-terminal may show a more dramatic reduction.

As proof of principle, we examined the ACINUS protein, whose gene is disrupted by T-DNA insertion. As expected, the N-terminal peptides before the T-DNA insertion in the *acinus pinin* double mutant show a 2-fold reduction relative to the same peptides in the wild type, but the C-terminal peptides after the T-DNA insertion show a dramatic reduction to undetectable levels. This quantification is consistent with our previous SILIA-PRM targeted quantification of the peptides showing that the N-terminal is reduced, and the C-terminal is not expressed in the double mutant as a null allele (10,16).

Following the analysis, we categorized each IIR-affected protein into six categories.

Category I includes proteins for which IIR has no effect on the transcript or protein levels. Category I contains 134 proteins (71.3%). Despite the intron retention, both N-terminal and C-terminal peptides show similar abundance to WT, as shown by the example of OM47 and SFR6 (Fig. 4A and Fig. 4C). Of these introns, 57 (42.5%) of these retained introns are known pts introns, 61 (45.5%) of them are known non-pts introns, and 16 of them are unknown.

Category II contains proteins for which the IIR does not change the transcript level, but significantly reduces the protein level. Category II includes 27 (14.4%) proteins. As indicated by examples of SGR6 and PPRT1 (Fig. 4A and Fig. 4D). In addition, 9 (33.3%) of these genes contain a pts intron, 14 (51.8%) are non-pts, and 4 are unknown.

Category III contains proteins for which IIR reduces both the transcript and protein levels. This category includes 22 genes (11.7%) and shows that both N-terminal and C-terminal peptides are down-regulated to a similar extent in *acinus pinin* mutants, such as LETM2 and BFA3 (Fig 3A and Fig 3E). Notably, these retained introns contain a premature stop codon. Of these genes 11 are shown to be pts introns, 10 are non-pts introns, and 1 is unknown. These transcripts may be exported out of the nucleus and degraded through the NMD pathway.

Category IV shows the cases where IIR increases the transcript level but does not affect the protein level. This category includes three genes such as by PAE8 (Fig. 4A and 4F). Conversely category V contains proteins where IIR decreases the transcript level but does not change in protein level. Category V contains a single protein, IDE2 (Fig. 4A and 4G).

Category VI contains proteins that are expected to be truncated based on the peptide analysis (Fig. 4H). We were able to detect one protein, GLOBAL TRANSCRIPTION FACTOR GROUP E8(GTE8), which encodes a kinase-like protein with similarities to yeast BDF1 and human RING3 protein, that produced a truncated protein. The N-terminal peptide before the retention intron shows no difference between WT and double mutant, while the C-terminal peptide shows a striking reduction, indicating that truncated proteins are produced. Notably, the intron of this GTE8 is highly retained as visualized by IGV (Fig. 4H), but the transcript level was similar to WT, showing that the transcript can escape the NMD and be translated.

We also examined the decreased intron retention events in the *acinus pnn* double mutants. Of the 31 genes with decreased retention, 22 were quantified by our TMT set. For these genes, the protein level remains unchanged in the *acinus pnn* mutant (Supplemental Figure 4).

Higher GC Frequency, longer introns and less optimal consensus sites observed in retained introns

To understand why only one or very few introns of the many introns in the same genes are retained in the *acinus pnn* mutants, we investigated the characteristics by comparing the GC content, intron length and splice consensus of retained introns and non-retained introns in the 225 genes affected by ACINUS and PININ loss. Higher GC content, weaker sites in retained introns have been observed in several mammalian studies (17,18). Consistent with these previous publications, we found that the GC content of retained introns is significantly higher than that of

non-retained introns, with a mean value (34.29% vs 33.33% GC content) and a p-value of 0.018. Similarly, we found that retained introns are slightly longer than the non-retained introns, with a mean value (153 vs 204 base pairs) and a p-value of 0.016. Next, we investigated whether retained introns have weaker splice sites. We examined the nucleotide frequency characteristics at 5' and 3' splice sites for spliced introns (top), retained introns (middle), and highly retained introns (bottom). The frequency of 'AG' nucleotides before the consensus 5'GT splice sites is lower in highly retained introns, while the frequency of "TT" nucleotides" before the consensus 3'AG splice sites is higher in highly retained introns, consistent with the general agreement that introns undergoing retention are more likely to have weak splice site reference. Taken together, these features of higher GC content, longer introns and weaker splice sites may collectively contribute to intron retention.

Phenotype prediction and validation guided using the proteomics data

Next, we use our proteomic data to predict the phenotype. Our Go analysis of up-regulated proteins in the *acinus pinin* mutant shows that the biological process of anthocyanin-containing compound biosynthetic processes is significantly enriched (Fig. 3C). As shown in Fig. 6A, several proteins involved in this process are significantly up regulated in the mutant. Therefore, we predict that the *acinus pinin* mutant will have a higher anthocyanin content. Indeed, when we grow both WT and *acinus pnn* on the same ½ MS plates, the *acinus pnn* double mutants have a clear accumulation of anthocyanin pigments under the cotyledons, making the cotyledons purple in color (Fig. 6B).

In addition, we observed that several proteins predicted to cause pale green, reduced growth are significantly reduced, such as SVR9L1 (19), LETM2 (20), NPL1-2 (21), KEA1 (22), BE3 (23), LDA (24), NHX5 (25) and SAMC1 (26). Except for BE3, all others have significantly increased intron retention in the mutants, leading to significant reduction in the protein levels (Fig. 7A and 7B). Together, these reduced proteins may explain the pale green and reduced growth in the double mutant (Fig. 7C). On the other hand, KEA1, CUALPHA2, JMJ30 and TAR2 have been shown to regulate primary root elongation, and reduction of these proteins may lead to the short root phenotype. These reduced proteins may help explain the short root phenotype observed in the *acinus pnn* double mutant.

Together, based on our proteomics data, we concluded that the pleiotropic effects of *acinus pnn* mutants are due to accumulated changes in protein levels that collectively contribute to the phenotype.

Discussion

Weak to modest correlation between altered gene expression and altered protein abundance.

To investigate whether altered gene expression correlates with altered protein abundance, we examined the contribution of changes in mRNA abundance to protein abundance using the *acinus pnn* mutant, which has altered gene expression and alternative splicing. We observed changes in gene expression for 1553 genes in the *acinus pnn* knockout and obtained fold change information for 811 proteins. Using 11,491 genes for which protein quantification was available for analysis, a Pearson correlation coefficient of 0.63 was observed with a p-value < 2.2e-16 (see Supplementary Figure 5), indicating that a significant proportion of the variation in protein abundance is primarily due to changes in mRNA levels. However, when analyzed separately for up- and down-regulated genes (fold cutoff=2), the correlation becomes weaker, ranging from 0.35 to 0.39. This suggests that for these genes, the protein fold changes are weakly correlated with the transcript changes. This low correlation may be due to the fact that some altered transcripts are not associated with changes in protein levels, while some genes have much higher protein upregulation than the transcript changes, such as PR genes.

Complex contribution of alternative splicing to protein abundance and diversity

It is currently believed that some intron-retained transcripts are subject to post-transcriptional splicing (PTS) regulation (9). Transcripts subjected to post translation splicing remain tethered and retained in the nucleus, and have splicing triggered by environmental and developmental cues (8,9). The PTS may or may not lead to a reduction in cytosolic RNA levels, thereby affecting protein levels in unpredictable ways. Alternatively, transcripts with intron retention may be transported to the cytosol and translated, but they may trigger nonsense-mediated decay (NMD), resulting in reduced protein abundance. In addition, transcripts may escape NMD, but the presence of a premature stop codon could result in a truncated protein (6).

Our analysis revealed that 71.3% of the intron-retained transcripts (IIR) do not show reduced transcript levels or reduced protein abundance. Of these, 42.5% were annotated as PTS introns. Furthermore, 14.4% of IIRs do not show changes in transcript levels, but have significantly reduced protein abundance, with 33.3% annotated as PTS introns. These results suggest that PTS introns may or may not affect overall transcript levels or protein abundance. On the other hand, we observe that 11.7% of the IIRs cause both reduced transcript and protein levels. Of these 22 genes, 11 are shown to be PTS introns, but it appears that they may be exported from the nucleus and subsequently degraded by NMD.

Intron retention could increase the diversity of the proteome. By escaping the NMD pathway, these transcripts are translated into truncated proteins. In our study, we detected only one clearly truncated protein, GTE8, which encodes a kinase-like protein similar to yeast BDF1 and human RING3 protein.

Despite our significant coverage of the proteome, we are aware of the potential limitations of our detection methods. For example, these IIRs may produce a truncated version or have significant effects at the protein level, but the intron retention retained in the double mutant is only a small percentage of these transcripts and therefore the truncated version or reduced protein level is averaged out. If the proportion of intron-retained versions is higher in the transcript, then the effects can be more readily detected. This may indeed be the case, for example, in some stress conditions. Indeed, using data mining, we have shown that the intron retention events in the *acinus pinin* double mutants can be replicated in the wild type under various stress conditions.

Alternatively, direct detection of the peptides located at the junction or from introns may provide a better understanding of the translation of intron retained transcripts. In our current study, we have used tryptic digestion to achieve the highest coverage to test for both transcriptional and alternative splicing effects. However, it has been shown that intron junctions are enriched in lysines, so they are cleaved, and the junctions are not easily detected, whereas chymotrypsin is recommended, but chymotrypsin has a much lower coverage than trypsin (27). On the other hand, a stop codon is often observed just after the junctions (28). Therefore, direct detection of these junctions and intron-encoded peptides is challenging. Better pipelines need to be developed to overcome these challenges.

Prediction from genotype to phenotype

Previously, we found that the hypersensitivity to ABA and the late flowering phenotype could be explained by altered transcripts in the ABA response and the FLC gene (10). Our new proteomic data support the notion that collective changes in the proteins may be the main contributors to the phenotypes. We show cases of collective protein changes that predict the higher anthocyanin and pale green and growth retardation phenotypes in our mutants, and indeed they are consistent with the phenotype observed in our double mutants. On the other hand, our data suggest that intron retention is a regulatory mechanism in plants in response to stress, and/or that ACINUS and PNN can be negatively regulated under these conditions, resulting in altered alternative splicing that affects protein expression and phenotype, so it will be interesting in the future to test the mutants under more stress conditions.

Methods

Plant materials and growth conditions

Arabidopsis thaliana (wild type and mutant) plants that were used in this study were Columbia-0 (WT) and *acinus-2 pinin-1* (ap) (10). Seeds were sterilized and stratified for 3 days. For tissues intended for the TMT experiment, seeds were sown on ½ MS agar and grown in a constant light chamber at 22 C for 9 days. For pale green and anthocyanin analyses, seedlings were 7 and 11 days old respectively. For root length assays. Seeds were sown on ½ MS agar and grown vertically in a constant light chamber at 22 C for 7 days. ImageJ was used to measure the root length as traced from the root node to the primary root tip. In total, the root length of WT (n=85) and *acinus pinin* (n=137) was measured.

TMT11 Multiplexing Samples preparation

100mg of tissue from 9-day-old of wild type and *acinus pinin* seedlings (5 biological replicates each) were harvested, flash frozen, and ground. Proteins were extracted as described in (29) with a slight modification of extraction buffer Y (100mM Tris-HCl, pH8.0, 2% (weight/vol) SDS, 5mM EGTA, 10mM EDTA, 1mM PMSF, 2x protease inhibitor (Roche), followed by reduction with TCEP and alkylation before tryptic digestion, then followed by desalting with Sep-pak columns (Waters).

TMT (11-plex) reagents (ThermoFisher, Cat#A34808) were reconstituted in anhydrous acetonitrile (0.01mg/uL). Approximately 200 µg of peptides from each sample were reconstituted in 66 uL of 50 mM HEPES buffer. Of These, 60 µL (182 µg) were mixed with 40 µL TMT reagents (TMT126 to 130N), and the remaining 6 µL (18.2 µg) were pooled and mixed with 40 µL TMT

reagent 131C. Samples were labeled for 1 h at 25°C with 700 rpm shaking and quenched with 5 µL of 1M Tris (pH8). Peptides were acidified by adding 45 µL of 10% FA in 10% acetonitrile. The final concentration of the reaction consisted of TMT at 4 µg/µL, peptides at 2 µg/µL, and acetonitrile at 40% (v/v) similar to (11). The specific labeling design of the TMT reagents is illustrated in Figure 2 A. One-eighth of these 11 labeled samples were then combined and subjected to extensive high pH reverse HPLC fractionations (29).

Fractionation and mass spectrometry analysis

High pH reverse-phase chromatography was performed on the Vanquish Flex system (ThermoFisher) equipped with a 4.6×150-mm Gemini 5µm C18 column for TMT-labeled peptides. Peptides were loaded onto the column in 25 µL of buffer A (20 mM ammonium formate, pH 10). Buffer B consisted of buffer A with 90% (vol/vol) acetonitrile in water. Separation was performed using the following gradient at a flow rate of 0.5 mL/min: Column equilibration was 10 minutes of 1%B. 1-10% B gradient for 20 minutes, 10-10% B gradient for 20 minutes, 10-50% gradient for 80 minutes, 50-70% gradient for 5 minutes, 70-100% gradient for 10 minutes,100% gradient for 30 minutes. The flow rate was 0.5 mL/min and fractions were collected for every minute. Fractions T49-T110 were collected and dried. Fractions T61-T99 were injected individually, while the remaining fractions were pooled. A total of 50 injections were measured on an Orbitrap Eclipse equipped with an Easy LC 1200 UPLC liquid chromatography system (Thermo Fisher). Peptides were first trapped in a trapping column (Acclaim PepMap 100 C18 HPLC, 75 µm particle size, 2 cm bed length) and then separated using an analytical column AUR3-25075C18, 25CM Aurora Series Emitter Column (25 cm x 75 µm, 1.7 µm C18) (IonOpticks). The flow rate was 300 nL/min, and a 135-min gradient was used. Peptides were eluted by a gradient of 3 to 6 solvent B (80 % acetonitrile, 0.1 % formic acid) for 1 min, 6 to 32% solvent B for 105 min and from 32 to 44 % solvent B for 15 min, followed by a wash at 90 % solvent B for 14 min.

The TMT real-time library search was similarly configured to Furtwangler et al (30). Briefly, on an Orbitrap Eclipse Tribrid Mass Spectrometer, the precursor scan was performed on the Orbitrap, scan range mass to charge ratio (m/z) 400 to 1600 (resolution 120,000; RF lens (%) 30). Intensity properties were set to filter with a threshold of 5.0E3. Theoretical precursor envelope fit properties were 50% and fit window was 0.7 (m/z). Charge state properties were 2-6. Dynamic exclusion was to be excluded after 1 time for 45 seconds. Data-Dependent MS2 was performed in the ion trap detector with an isolation window (m/z) 0.7, collision energy was CID at 35%. The maximum

injection time was 35ms. For real-time library search, the modified TAIR10 database (as described in TMT LC-MS-MS Data Analysis methods) was used with carbamidomethyl and TMT11plex as the static modifications and with oxidation as variable modification. MS3 scan was enabled to use synchronous precursor selection (SPS) with the number of precursors set to 10, the isolation window set to 0.7 and the HCD collision energy set to 65%. The Orbitrap was set to a resolution of 50000, and a scan range of 100-500 (m/z) with a maximum injection time of 200ms. The MS3 cycle time was 2.5 seconds.

TMT LC-MS-MS Data Analysis

For Eclipse MS data, raw files were searched using MaxQuant v2.4.2.0. In the global parameters, the type was set to MS3 ion reporter mode using the correction factors provided by the TMT reagents. No normalization setting was set. The reference protein database used was a modified version of the Arabidopsis Thaliana TAIR 10 database, downloaded from TAIR: "TAIR10_pep_20101214.fa". The reference protein database included the addition of in-silico translated proteins from genes with increased intron retention in *acinus pinin.* This database includes the incorporation of the retained intron for the predicted protein. All other parameters remained at their default.

For protein level analysis: MaxQuant outputs proteinGroups.txt and evidence.txt were used with R package MStatsTMT (31). The parameters for MSstatsTMT's proteinSummarization function: method was set to msstats, global normalization was set to true, reference normalization was set to true, and remove normalization channel was set to true. Proteins with missing p-values were filtered out. If an ambiguous protein group existed with different gene loci, they were discarded. If multiple protein groups existed for a single gene locus, the protein group with the most peptides quantified was kept. Protein groups with a p-value less than 0.05 and a fold change greater than 1 or less than -1 were considered as differentially abundant proteins. See Supplementary Figure 1A. Integration of transcriptomic data: RNA-seq data were obtained from Bi et. al (10). Using R we merged this dataset with the TMT dataset using the GeneID column.

For peptide level analysis: MaxQuant outputs peptides.txt were used with Perseus v2.0.11.0 and R. In Perseus, the data were log2 transformed and filtered for at least 5 replicates per condition (wild type, *acinus pinin*). In R the data were subjected to global median normalization (Supplementary Figure 2B). The resulting data were re-imported into Perseus to impute missing values from a normal distribution with a width of 0.3 and downshifted 1.8 from the values of the

total matrix. The data were re-imported into R to remove non-unique peptides. The approximate position of an intron on a protein was calculated by constructing the transcript with the affected intron and translating it into a protein in-silico. The length of the resulting protein was used for the “before” and “after” categorization of peptides, see Supplementary Figure 2A. For abundance processing, for each measured peptide, each channel was then divided by the median log2 abundance of the wild type samples, Supplementary Figure 3A, B.

Data Mining Intron retention of abiotic and biotic stresses

Information on introns significantly affected by various abiotic and biotic treatments was collected from materials curated by Jia et al (9). A custom R script was used to find all introns that overlapped with our increased retention set and the increased retained introns in various abiotic and biotic studies. The intron retention data from each treatment for relevant introns were downloaded from the Plant Intron Splicing Efficiency Database (PISE) (7).

Genomic Feature Analysis

The following R packages were used for GC content and splice site motif analysis: GenomicFeatures (32), BSgenome, GenomicRanges (32), and Biostrings. Briefly, a BSGenome object was created using the Tair10 reference genome. Annotations were added as a Tx.Db object. For each gene affected by increased intron retention in *acinus pinin* genomic ranges were retrieved for each intron. GC content was calculated for each intron by retrieving the sequence of each intron genomic range. To analyze sequence motifs the flanking genomic ranges of each intron were obtained. Windows of 12 bases were retrieved and placed in a position matrix to generate the sequence motif using ggseqlogo (33).

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository. All other data are available from the corresponding author on reasonable request.

Acknowledgements

We thank Dr. Jixian Zhai and Dr. Jinbu Jia for providing the Arabidopsis thaliana increased intron retention datasets for various biotic and abiotic conditions. This work was funded by the National

Institutes of Health grants R01GM135706 and S10OD030441 to S.-L.X. and diversity supplement to support A.V.R, and by the Carnegie Endowment Fund to the Carnegie Mass Spectrometry Facility.

Author contributions – S.-L.X. conceptualization; A.V.R. and T.S.G. TMT sample preparation. A.V.R. data analysis and figures. T.S.G., R.S., D.B., technical support. A.V.R. and S.-L.X. manuscript writing.

Reference

1. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457-463.
2. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, **40**, 1413-1415.
3. Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A.S.N., Byrne, L.J., Wilson, C.M. and Syed, N.H. (2019) Alternative Splicing and Protein Diversity: Plants Versus Animals. *Front Plant Sci*, **10**, 708.
4. Boutz, P.L., Bhutkar, A. and Sharp, P.A. (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev*, **29**, 63-80.
5. Jacob, A.G. and Smith, C.W.J. (2017) Intron retention as a component of regulated gene expression programs. *Hum Genet*, **136**, 1043-1057.
6. Wright, C.J., Smith, C.W.J. and Jiggins, C.D. (2022) Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet*, **23**, 697-710.
7. Zhang, H., Jia, J. and Zhai, J. (2023) Plant Intron-Splicing Efficiency Database (PISE): exploring splicing of approximately 1,650,000 introns in Arabidopsis, maize, rice, and soybean from approximately 57,000 public RNA-seq libraries. *Sci China Life Sci*, **66**, 602-611.
8. Yan, Y., Luo, H., Qin, Y., Yan, T., Jia, J., Hou, Y., Liu, Z., Zhai, J., Long, Y., Deng, X. *et al.* (2024) Light controls mesophyll-specific post-transcriptional splicing of photoregulatory genes by AtPRMT5. *Proc Natl Acad Sci U S A*, **121**, e2317408121.
9. Jia, J., Long, Y., Zhang, H., Li, Z., Liu, Z., Zhao, Y., Lu, D., Jin, X., Deng, X., Xia, R. *et al.* (2020) Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat Plants*, **6**, 780-788.
10. Bi, Y., Deng, Z., Ni, W., Shrestha, R., Savage, D., Hartwig, T., Patil, S., Hong, S.H., Zhang, Z., Oses-Prieto, J.A. *et al.* (2021) Arabidopsis ACINUS is O-glycosylated and regulates transcription and alternative splicing of regulators of reproductive transitions. *Nat Commun*, **12**, 945.
11. Zecha, J., Satpathy, S., Kanashova, T., Avanessian, S.C., Kane, M.H., Clauser, K.R., Mertins, P., Carr, S.A. and Kuster, B. (2019) TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics*, **18**, 1468-1478.
12. Schweppe, D.K., Eng, J.K., Yu, Q., Bailey, D., Rad, R., Navarrete-Perea, J., Huttlin, E.L., Erickson, B.K., Paulo, J.A. and Gygi, S.P. (2020) Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J Proteome Res*, **19**, 2026-2034.

13. Lu, H.P., Gao, Q., Han, J.P., Guo, X.H., Wang, Q., Altosaar, I., Barberon, M., Liu, J.X., Gatehouse, A.M.R. and Shu, Q.Y. (2022) An ABA-serotonin module regulates root suberization and salinity tolerance. *New Phytol*, **236**, 958-973.
14. Baker, P.R. and Chalkley, R.J. (2014) MS-viewer: a web-based spectral viewer for proteomics results. *Mol Cell Proteomics*, **13**, 1392-1396.
15. Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M. and Blencowe, B.J. (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*, **24**, 1774-1786.
16. Reyes, A.V., Shrestha, R., Baker, P.R., Chalkley, R.J. and Xu, S.-L. (2022) Application of Parallel Reaction Monitoring in 15N Labeled Samples for Quantification. *Front. Plant Sci.*, **13**, 832585.
17. Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B. *et al.* (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep*, **1**, 543-556.
18. Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M.C., Shini, S., Lieschke, G.J., Wong, J.J. and Rasko, J.E.J. (2017) Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol*, **18**, 216.
19. Zheng, M., Liu, X., Liang, S., Fu, S., Qi, Y., Zhao, J., Shao, J., An, L. and Yu, F. (2016) Chloroplast Translation Initiation Factors Regulate Leaf Variegation and Development. *Plant Physiol*, **172**, 1117-1130.
20. Zhang, B., Carrie, C., Ivanova, A., Narsai, R., Murcha, M.W., Duncan, O., Wang, Y., Law, S.R., Albrecht, V., Pogson, B. *et al.* (2012) LETM proteins play a role in the accumulation of mitochondrially encoded proteins in Arabidopsis thaliana and AtLETM2 displays parent of origin effects. *J Biol Chem*, **287**, 41757-41773.
21. Feldman, M.J., Poirier, B.C. and Lange, B.M. (2015) Misexpression of the Niemann-Pick disease type C1 (NPC1)-like protein in Arabidopsis causes sphingolipid accumulation and reproductive defects. *Planta*, **242**, 921-933.
22. Aranda-Sicilia, M.N., Aboukila, A., Armbruster, U., Cagnac, O., Schumann, T., Kunz, H.H., Jahns, P., Rodriguez-Rosales, M.P., Sze, H. and Venema, K. (2016) Envelope K+/H+ Antiporters AtKEA1 and AtKEA2 Function in Plastid Development. *Plant Physiol*, **172**, 441-449.
23. Dumez, S., Wattebled, F., Dauvillee, D., Delvalle, D., Planchot, V., Ball, S.G. and D'Hulst, C. (2006) Mutants of Arabidopsis lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. *Plant Cell*, **18**, 2694-2709.
24. Delatte, T., Umhang, M., Trevisan, M., Eicke, S., Thorneycroft, D., Smith, S.M. and Zeeman, S.C. (2006) Evidence for distinct mechanisms of starch granule breakdown in plants. *J Biol Chem*, **281**, 12050-12059.
25. Wang, L., Wu, X., Liu, Y. and Qiu, Q.S. (2015) AtNHX5 and AtNHX6 Control Cellular K+ and pH Homeostasis in Arabidopsis: Three Conserved Acidic Residues Are Essential for K+ Transport. *PLoS One*, **10**, e0144716.
26. Bouvier, F., Linka, N., Isner, J.C., Mutterer, J., Weber, A.P. and Camara, B. (2006) Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *Plant Cell*, **18**, 3088-3105.
27. Sinitcyn, P., Richards, A.L., Weatheritt, R.J., Brademan, D.R., Marx, H., Shishkova, E., Meyer, J.G., Hebert, A.S., Westphall, M.S., Blencowe, B.J. *et al.* (2023) Global detection of human variants and isoforms by deep proteome sequencing. *Nat Biotechnol*, **41**, 1776-1786.
28. Petrillo, E. (2023) Do not panic: An intron-centric guide to alternative splicing. *Plant Cell*, **35**, 1752-1761.

29. Xu, S.-L., Chalkley, R.J., Maynard, J.C., Wang, W., Ni, W., Jiang, X., Shin, K., Cheng, L., Savage, D., Hühmer, A.F.R. *et al.* (2017) Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E1536-E1543.
30. Furtwangler, B., Uresin, N., Motamedchaboki, K., Huguet, R., Lopez-Ferrer, D., Zabrouskov, V., Porse, B.T. and Schoof, E.M. (2022) Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Mol Cell Proteomics*, **21**, 100219.
31. Huang, T., Choi, M., Tzouros, M., Golling, S., Pandya, N.J., Banfai, B., Dunkley, T. and Vitek, O. (2020) MSstatsTMT: Statistical Detection of Differentially Abundant Proteins in Experiments with Isobaric Labeling and Multiple Mixtures. *Mol Cell Proteomics*, **19**, 1706-1723.
32. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol*, **9**, e1003118.
33. Wagih, O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645-3647.

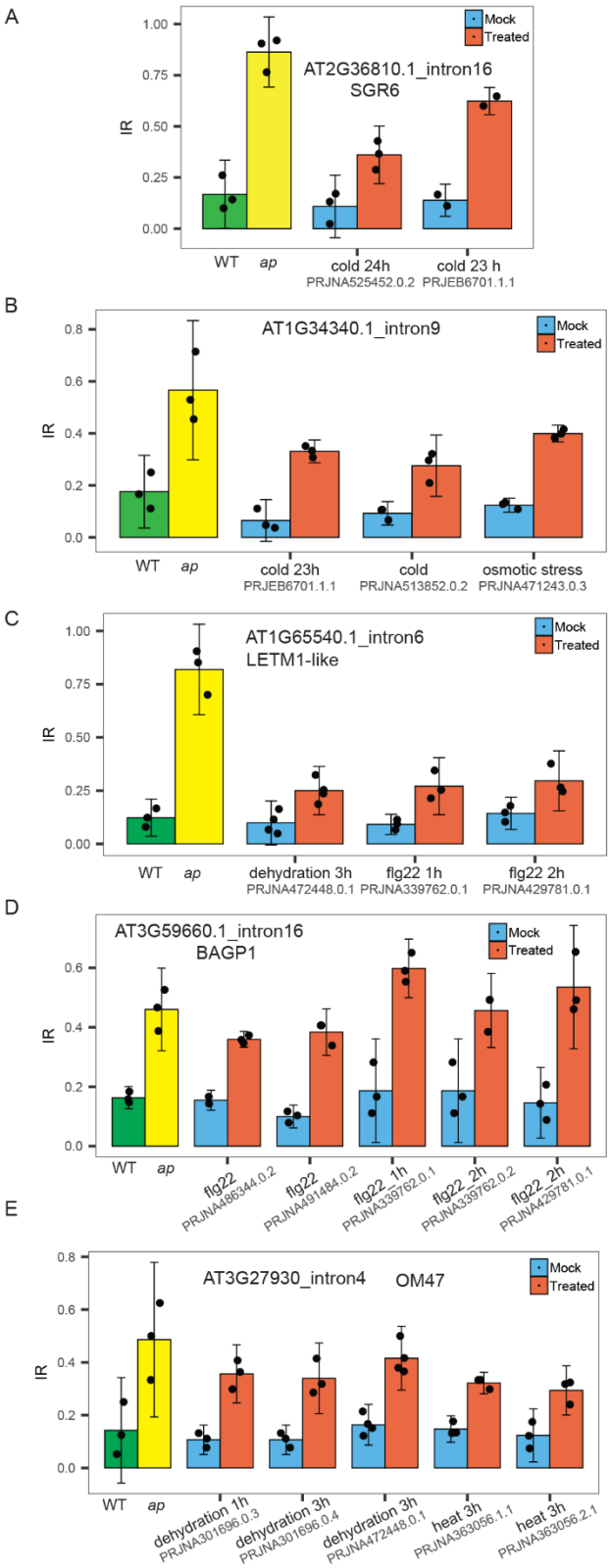
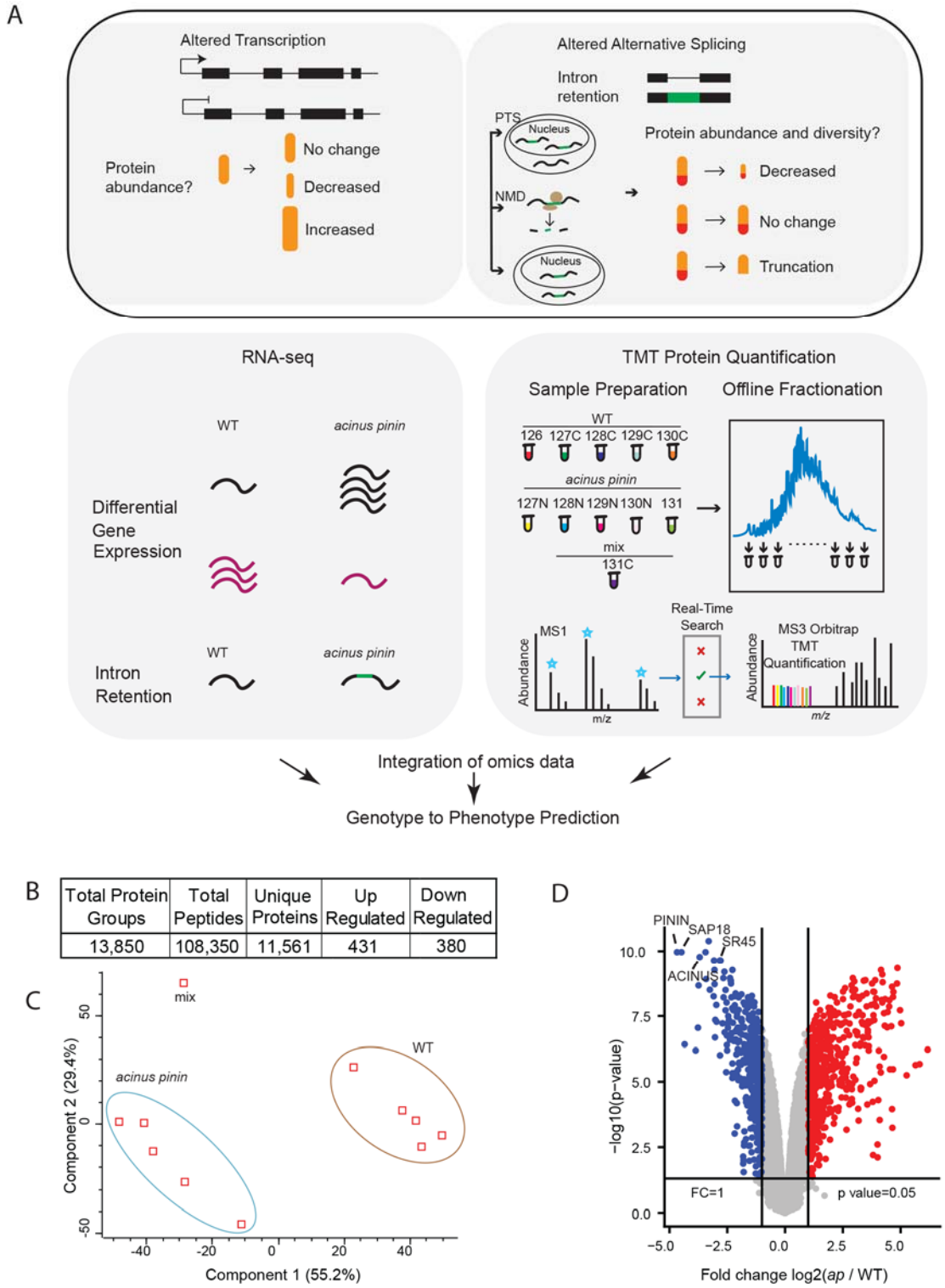

A
AT2G36810.1_intron16
SGR6
Mock
Treated
IR
1.00
0.75
0.50
0.25
0.00
WT
ap
cold 24h
PRJNA525452.0.2
cold 23 h
PRJEB6701.1.1
B
AT1G34340.1_intron9
Mock
Treated
IR
0.8
0.6
0.4
0.2
0.0
WT
ap
cold 23h
PRJEB6701.1.1
cold
PRJNA513852.0.2
osmotic stress
PRJNA471243.0.3
C
AT1G65540.1_intron6
LETM1-like
Mock
Treated
IR
1.00
0.75
0.50
0.25
0.00
WT
ap
dehydration 3h
PRJNA472448.0.1
flg22 1h
PRJNA339762.0.1
flg22 2h
PRJNA429781.0.1
D
AT3G59660.1_intron16
BAGP1
Mock
Treated
IR
0.6
0.4
0.2
0.0
WT
ap
flg22
PRJNA486344.0.2
flg22
PRJNA491484.0.2
flg22_1h
PRJNA339762.0.1
flg22_2h
PRJNA339762.0.2
flg22_2h
PRJNA429781.0.1
E
AT3G27930_intron4
OM47
Mock
Treated
IR
0.8
0.6
0.4
0.2
0.0
WT
ap
dehydration 1h
PRJNA301696.0.3
dehydration 3h
PRJNA301696.0.4
dehydration 3h
PRJNA472448.0.1
heat 3h
PRJNA363056.1.1
heat 3h
PRJNA363056.2.1

Figure 1: Intron retention events detected in the *acinus pinin* mutants are physiologically replicated in the wild-type plants under biotic or abiotic stress conditions. (A-E) shows bar graphs of six examples of introns retained in the double mutants that are similarly retained in the wild-type plants upon cold, osmotic, dehydration, flg 22, or heat treatment. Specific project names for each RNA-seq experiment are listed. Error bars represent the mean plus or minus the standard deviation.



Total Protein Groups	Total Peptides	Unique Proteins	Up Regulated	Down Regulated
13,850	108,350	11,561	431	380

Figure 2: An integrated workflow coupling RNA-seq with TMT quantitative proteomic analysis to study the contribution of alternative splicing to protein expression and diversity. (A) Experimental workflow. Both transcription and alternative splicing are perturbed in *acinus pinin* mutants, which can lead to different consequences as hypothesized in the examples. Altered transcripts can lead to increased, decreased, or unchanged protein

abundance. Transcripts with intron retention can be tethered by chromatin that is regulated by post-transcriptional splicing (PTS), or transported to the cytosol and undergo nonsense-mediated decay (NMD), or escape the NMD and be translated. The consequences can be reduced protein abundance, no change, or truncated proteins produced. TMT quantitative proteomics analysis is set up with 5 biological replicates for WT and *acinus pnn* mutants using TMT 11 plex reagents, fractionated and analyzed by LC-MS-MS with real-time search and SPS-MS3 acquisition. (B) Summary of TMT quantification. (C) PCA plot of the TMT experiment showing high data quality, with wild-type and *acinus pnn* clustering between biological replicates, distinct from each other and from the mix channel. (D) Volcano plot of TMT quantification showing 431 up-regulated proteins (red) and 380 down-regulated proteins (blue) using a fold cut-off of log 2=1 and a p-value of 0.05. Positive controls of significantly downregulated proteins such as ACINUS, PININ, SAP18 and SR45 are labeled.

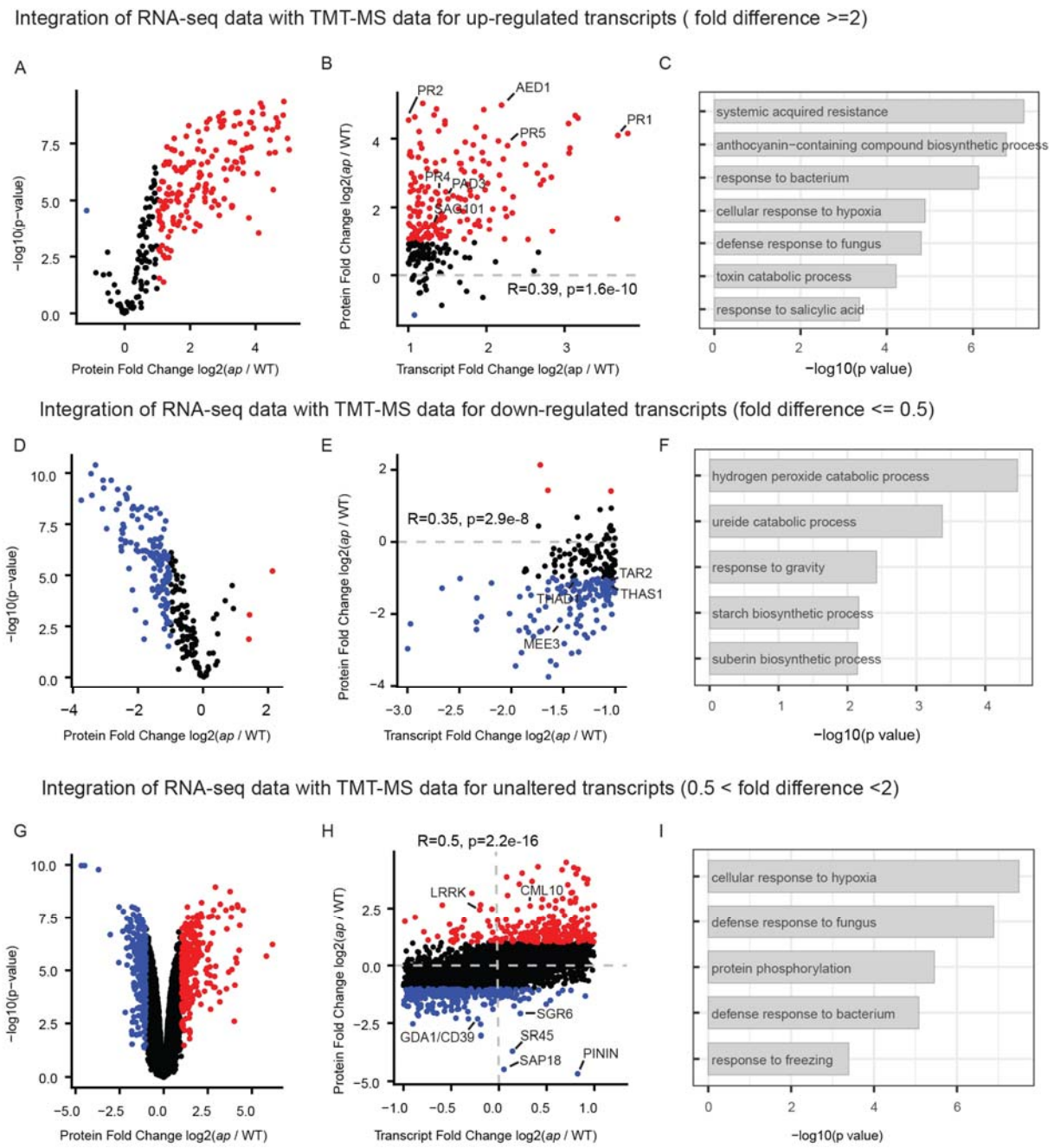
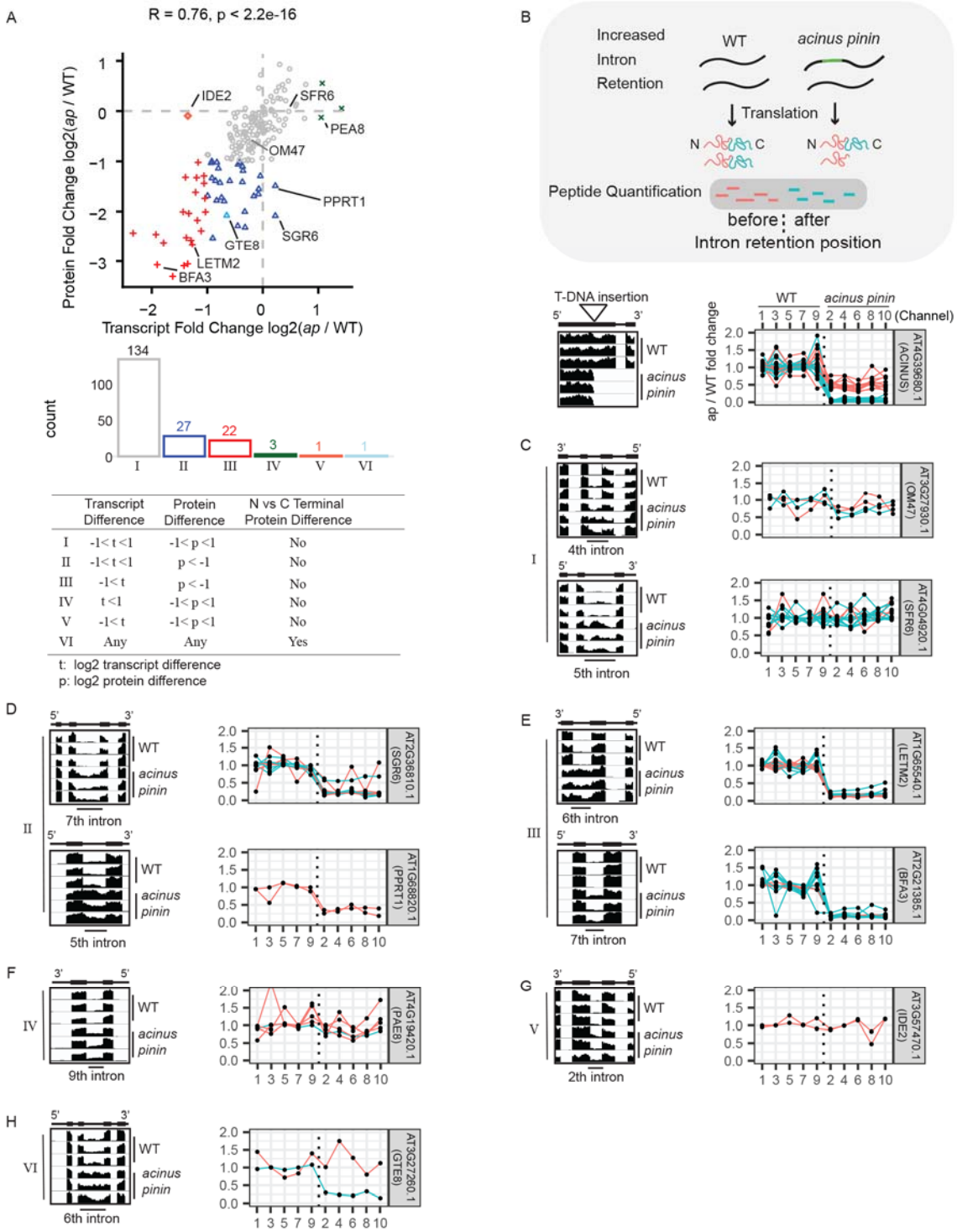


Figure 3: Comparison of transcriptomic and proteomic difference shows weak to moderate correlation. The 2-fold difference is used for both transcript and protein change cut-offs. (A, D, G). Volcano plots of abundance differences of quantified proteins corresponding to genes with increased transcript levels (A), or decreased transcript levels (D) or unchanged levels (G). (B, E, H) Scatter plots of protein fold changes to corresponding gene transcript changes for each category (increased, decreased, and unchanged transcript levels). Those colored in red represent transcripts that show significantly increased protein abundance in the *ap* mutant compared to the wild type, while those colored in blue represent significantly decreased protein abundance. (C) GO analysis of those genes in B that are up-regulated at

both the transcript and protein levels. (F) GO analysis of those genes in E that are downregulated at both the transcript and protein levels. (I) GO analysis of genes in H that show unchanged transcripts but have an increase at the protein level.



	Transcript Difference	Protein Difference	N vs C Terminal Protein Difference
I	-1< t <1	-1< p <1	No
II	-1< t <1	p < -1	No
III	-1< t	p < -1	No
IV	t <1	-1< p <1	No
V	-1< t	-1< p <1	No
VI	Any	Any	Yes

t: log2 transcript difference
p: log2 protein difference

Figure 4: Contribution of intron retention to protein expression and diversity. (A) Scatter plot integrating transcriptomic data with proteomic data of genes with increased intron retention in the *acinus pnn* mutant (top*)*. Bar plots contain the number of genes grouped into six different categories according to their change in transcript expression and protein abundance change in the *acinus pnn* mutant compared to wild type (middle*)*. A table defining how each gene fits into one of the six categories (bottom*)*. (B) Workflow to determine the contribution of IR to protein

expression and diversity. Quantified peptides before and after the intron position are shown in red and blue, respectively. For each graph, WT sample channels (1,3,5,7,9) are shown on the left and *acinus pinnin* mutant sample channels (2,4,6,8,10) are shown on the right. The example of peptide quantification for ACINUS is shown, where the red peptides before the T-DNA insertions show a reduction in the double mutant, but the C-terminals show dramatic reduction to the noise level. (C) Category I showing that IIR has no effect on the transcript or protein levels. (D) Category II showing that IIR does not change the transcript level but reduces the protein level. (E) Category III showing that IIR reduces both the transcript and protein levels. (F) Category IV showing that IIR increases the transcript level but does not affect the protein level. (G). Category V showing that IIR decreases transcription but no change in protein (H). Category V showing that IIR produces a truncated protein.

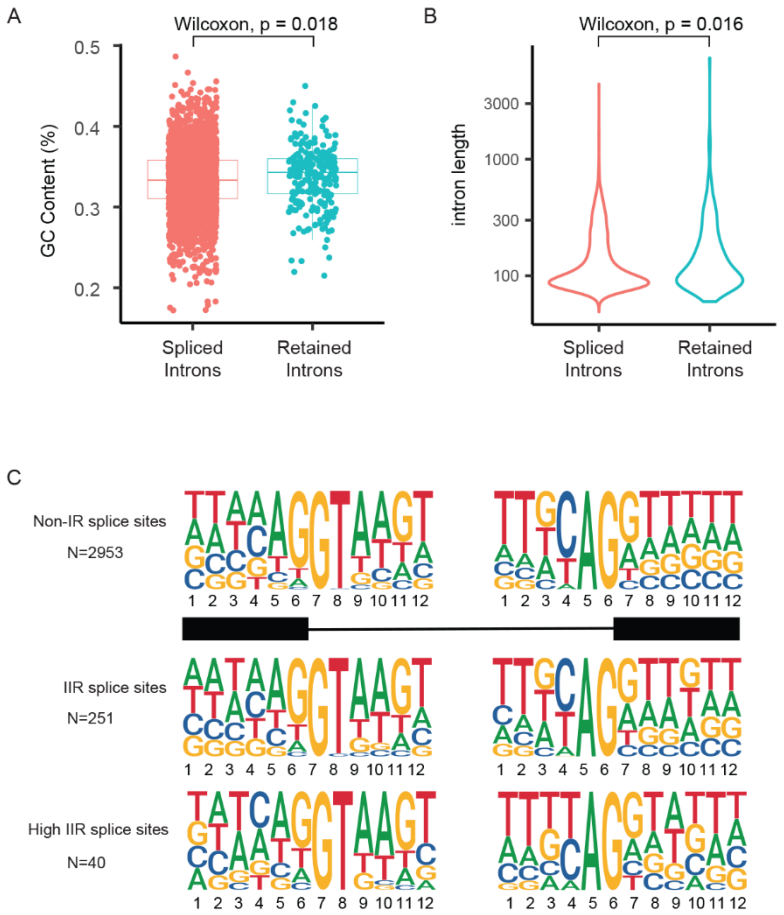


Figure 5: Characteristics of introns that are retained in the *acinus pinin* mutant. Introns from the affected genes with increased intron retention in the *ap* double mutants are retrieved for data analysis, with retained introns colored in blue and spliced introns colored in red. (A) Box plot showing that the GC consents of retained introns are significantly higher from that of spliced introns, with a mean value (33.33% vs 34.29%). (B) Violin plot showing that the length of retained introns is significantly longer than that of spliced introns. (C) Examination of the nucleotide frequency characteristics at 5' and 3' splice sites for spliced introns (top), retained intron (middle), highly retained intron (bottom). The frequency of 'AG' nucleotides before the consensus 5'GT splice sites is lower in highly retained introns, while the frequency of "TT" nucleotides" before the consensus 3'AG splice sites is higher in highly retained introns.

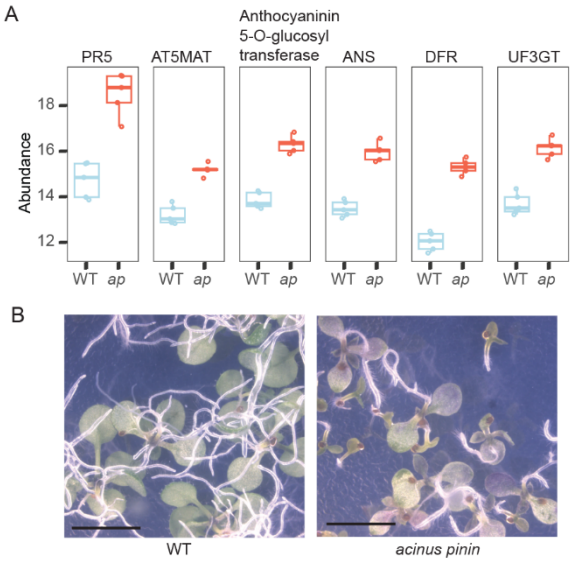
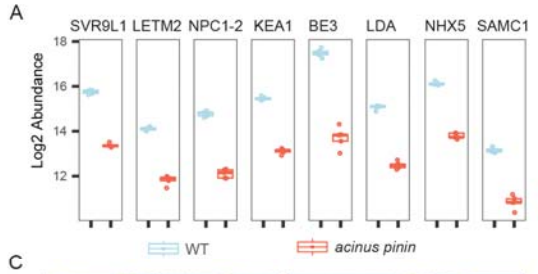
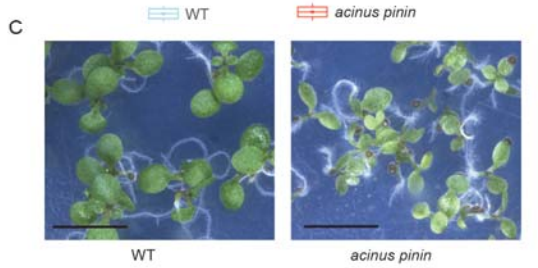


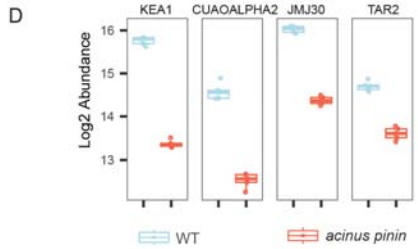
Figure 6: Upregulated proteins involved in anthocyanin biogenesis proteins are confirmed by increased anthocyanin in the *acinus pinin* mutant. (A) Box plot showing the increased abundance of a set of proteins involved in the biological process of anthocyanin biogenesis in *acinus pinin* (red) compared to wild type (blue). (B) Images taken from the abaxial side of cotyledons showing wild type and *acinus pinin* populations with increased anthocyanin in the *acinus pinin* mutants. Scale bar is 2mm.



B

Symbol	IIR	Phenotype	Reference
SVR9L1	yes	Double Mutant with SVR9 exhibit pale green color	Zheng et al., 2016
LETM2	yes	Hemizygous with LETM2 exhbit slow growth	Zhang et al., 2012
NPC1-2	yes	Double knockout with ATNPC1-1 leads to sevear dwarfism	Feldman et al., 2015
KEA1	yes	Double knockout with ATKEA2 leads to pale green color	Nieves et al., 2016
BE3	no	Double kockout with BE2 lead to pale green color and reduced size	Dumez et al., 2006
LDA	yes	Double kockout with atisa3-2 cause delayed growth	Delatte et al., 2006
NHX5	yes	Double kockout with NHX6 have reduced growth	Wang et al., 2015
SAMC1	yes	Single mutants have pale green leaf color and sevear growth retardation	Bouvier et al., 2006





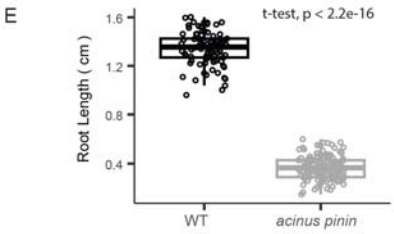
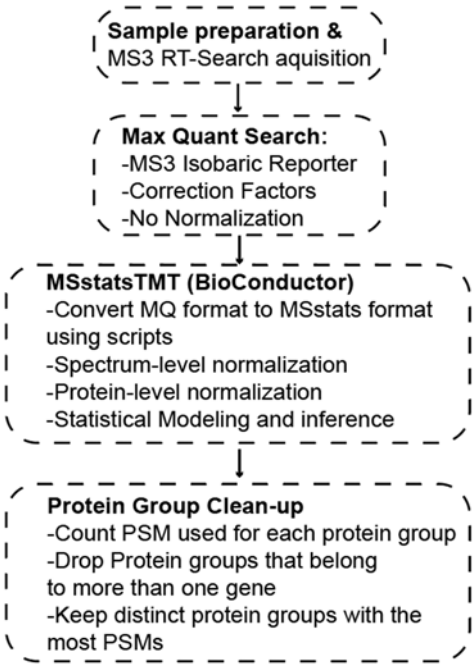
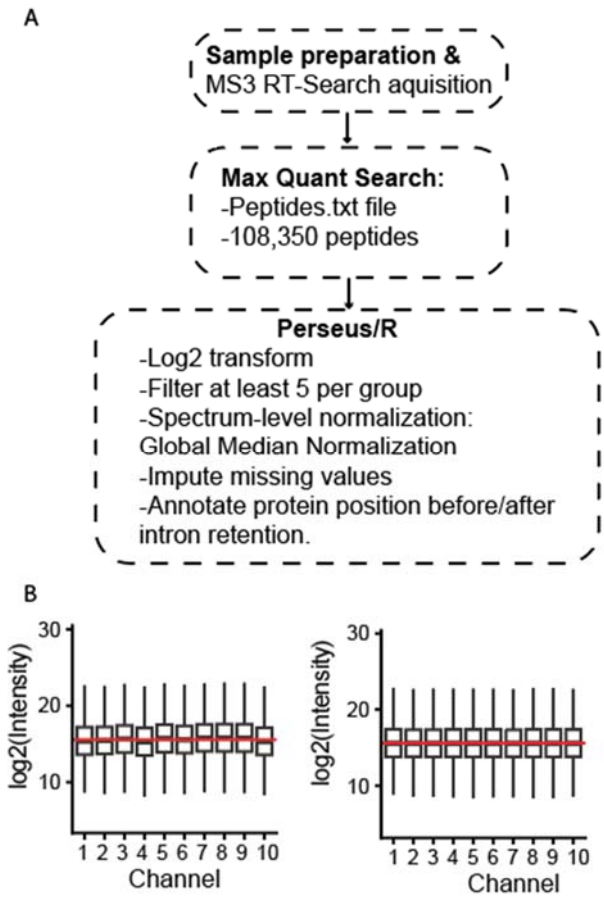


Figure 7: Protein changes correlate with phenotypic traits in the *acinus pinin* mutant.
(A, B, C) Proteins exhibiting significantly reduced protein levels (A) in the *acinus pnn* mutants are associated with pale green and/or reduced growth (B). The collective reduction of these proteins likely contributes to the observed pale green and reduced growth phenotype in the *acinus pnn* mutants (C). Scale bar = 2mm.
(D) Proteins displaying significantly reduced protein levels in the double mutants are implicated in primary root growth. The combined decrease in these proteins predict the double mutant would have a short root.
(E) Box plots depict root length of wild type (WT) and *acinus pnn* mutants. Data were analyzed by Student's t-test.

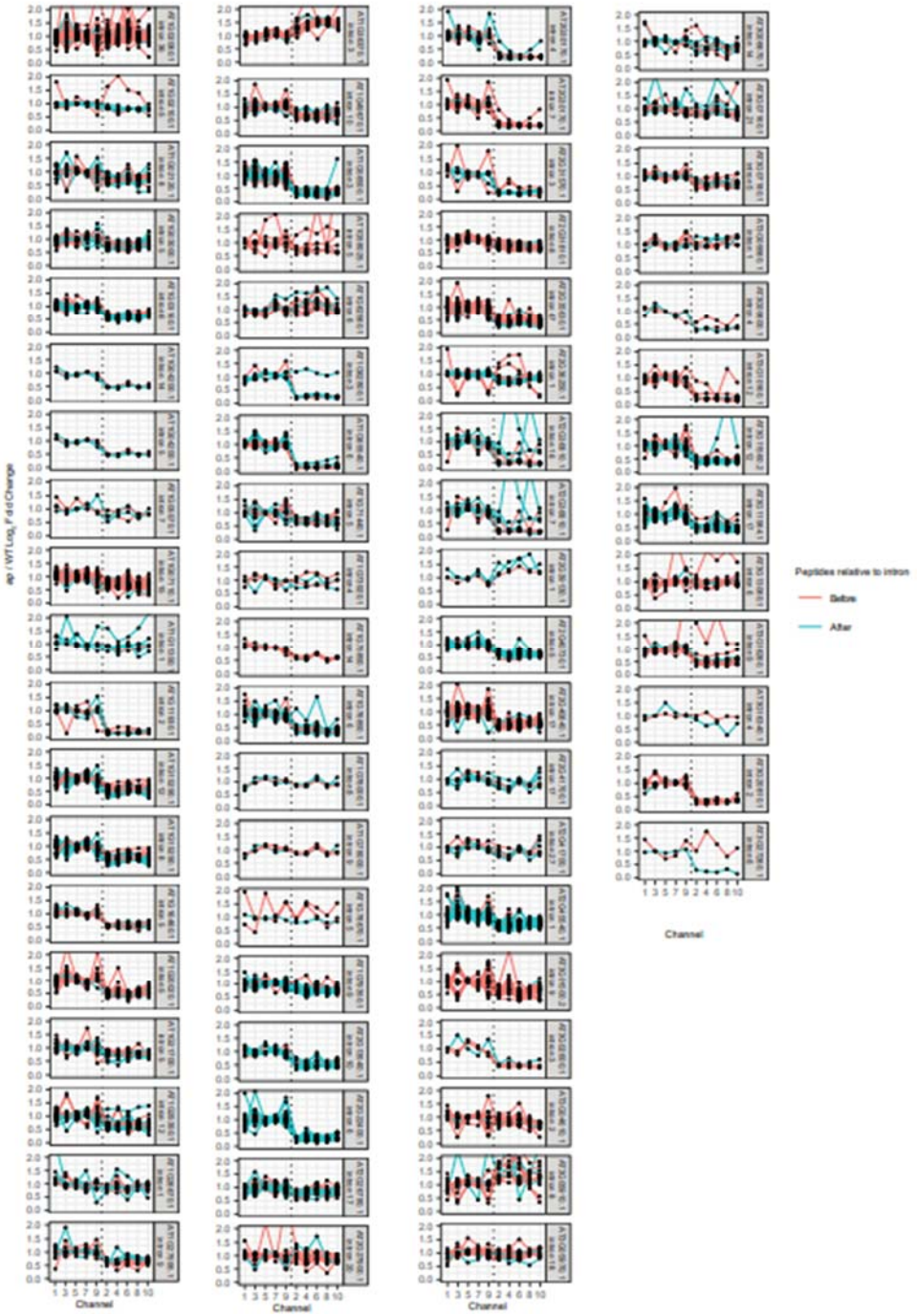


Supplementary Figure 1: Protein level processing workflow for TMT real time MS3 data. (A) MaxQuant search on TMT raw data. MSstatsTMT was used for normalization and statistical analysis. To obtain single quantification for each protein product we counted the number of peptide spectrum matches (PSM) and kept the protein group that is unique to one gene and has the most spectrum matches. This was done in Rstudio.

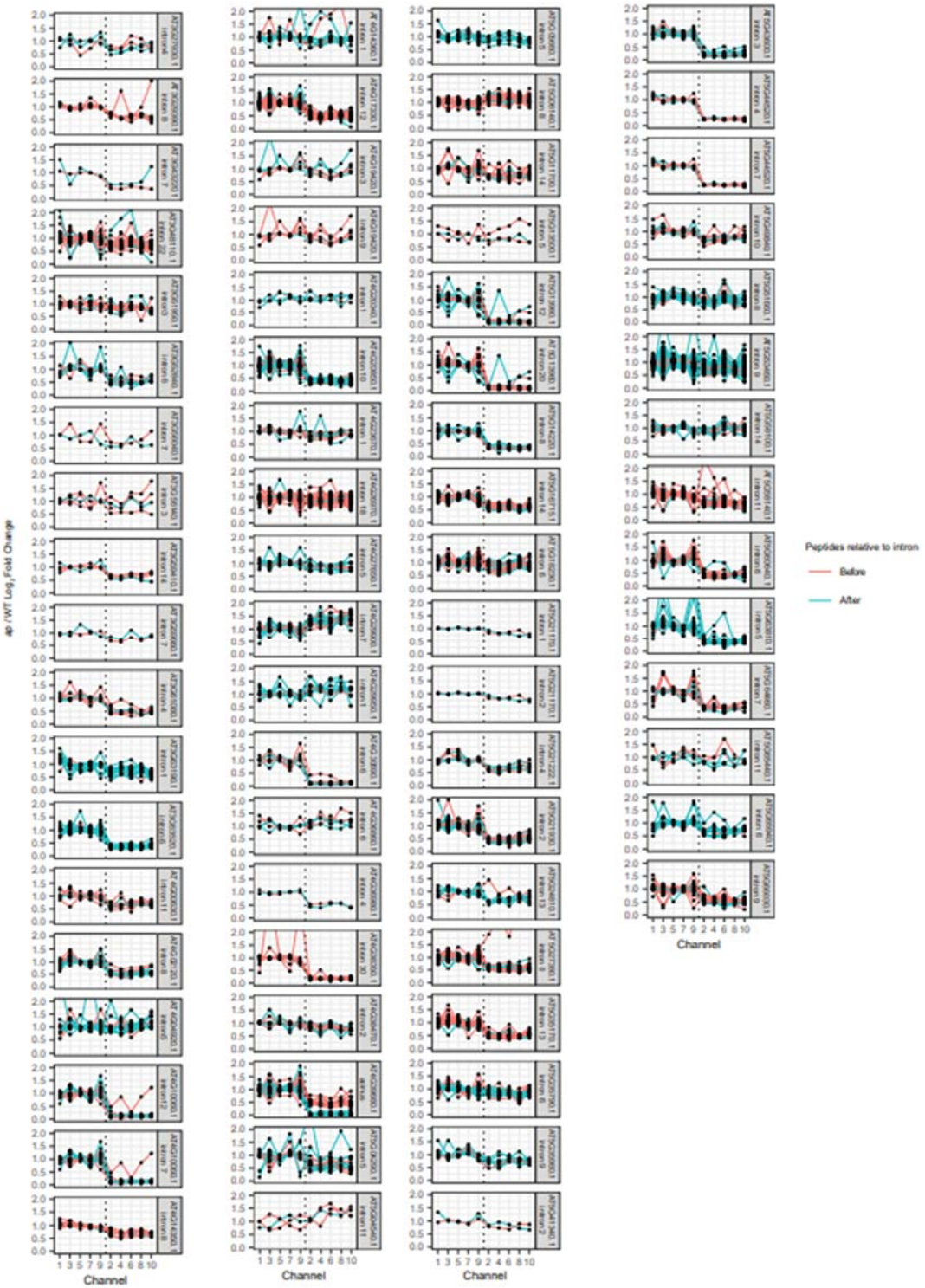


Supplementary Figure 2: Peptide level processing workflow for TMT Realtime MS3 data. (A) MaxQuant search on TMT raw data. The peptide output of MaxQuant was imported into Perseus where it was log transformed, filtered for at least 5 per condition (wildtype, acinus pinin) and missing values were imputed. Global median normalization was performed in R (B). Annotation of intron position was done in R.

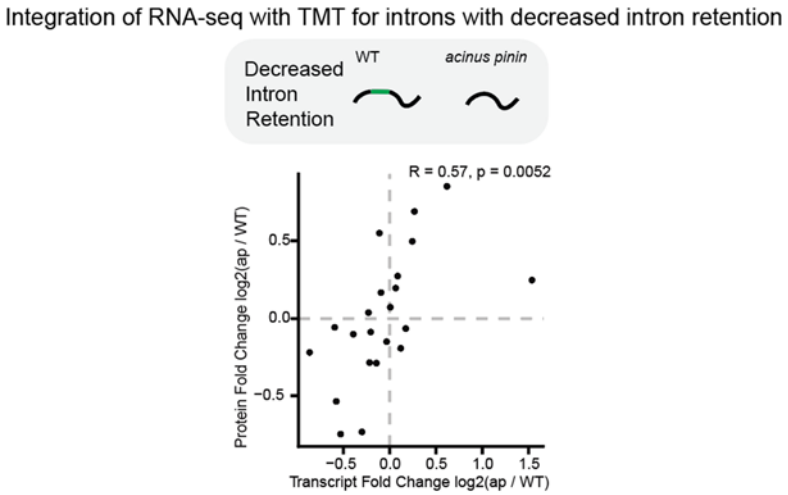
A



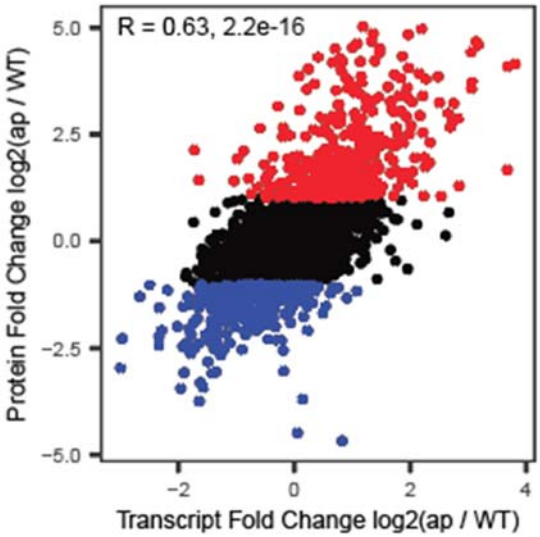
B



Supplementary Figure 3: Peptide level visualization of TMT data with respect to intron retention position. (A,B) line plots showing the relative abundance of each channel relative to the WT of each peptide quantified for the given protein.



Supplementary Figure 4: proteins with transcripts with decreased intron retention do not show increase of protein abundance. Scatter plot shows the relationship with transcripts that undergo decreased intron retention: their transcript expression and their protein abundance. Pearson correlation was conducted.



Supplementary Figure 5: Overall correlation of RNAseq to TMT proteomics is moderate. **(**A) Transcript fold change plotted with protein fold change. Red are significantly increased in the proteomic data set and blue are significantly decreased in the proteomic dataset.