Combining tRNA sequencing methods to characterize plant tRNA expression and post-transcriptional modification

Jessica M. Warren[1], Thalia Salinas-Giegé[2], Guillaume Hummel[2], Nicole L. Coots[1], Joshua M. Svendsen[1], Kristen C. Brown[1], Laurence Drouard, Daniel B. Sloan[1]

[1]Department of Biology, Colorado State University, Fort Collins, CO 80523-1878, USA
[2]Institut de biologie moléculaire des plantes-CNRS, Université de Strasbourg, F-67084 Strasbourg, France

Corresponding author: warrenj@rams.colostate.edu

ABSTRACT

Differences in tRNA expression have been implicated in a remarkable number of biological processes. There is growing evidence that tRNA genes can play dramatically different roles depending on both expression and post-transcriptional modification, yet sequencing tRNAs to measure abundance and detect modifications remains challenging. Their secondary structure and extensive post-transcriptional modifications interfere with RNA-seq library preparation methods and have limited the utility of high-throughput sequencing technologies. Here, we combine two modifications to standard RNA-seq methods by treating with the demethylating enzyme AlkB and ligating with tRNA-specific adapters in order to sequence tRNAs from four species of flowering plants, a group that has been shown to have some of the most extensive rates of post-transcriptional tRNA modifications. This protocol has the advantage of detecting full-length tRNAs and sequence variants that can be used to infer many post-transcriptional modifications. We used the resulting data to produce a modification index of almost all unique reference tRNAs in *Arabidopsis thaliana*, which exhibited many anciently conserved similarities with humans but also positions that appear to be "hot spots" for modifications in angiosperm tRNAs. We also found evidence based on northern blot analysis and droplet digital PCR that, even after demethylation treatment, tRNA-seq can produce highly biased estimates of absolute expression levels most likely due to biased reverse transcription. Nevertheless, the generation of full-length tRNA sequences with modification data is still promising for assessing differences in relative tRNA expression across treatments, tissues or subcellular fractions and help elucidate the functional roles of tRNA modifications.

INTRODUCTION

Despite their central function in cell physiology and the increasing interest in quantifying their expression, transfer RNAs (tRNAs) remain difficult to sequence using conventional RNA sequencing (RNA-seq) methods. tRNAs are poorly suited to high-throughput sequencing library preparation protocols for two reasons. First, tRNAs are extensively post-transcriptionally modified, and certain base modifications occurring at the Watson-Crick face stall or terminate reverse transcription (RT) through interference with base pairing or steric hindrance, effectively blocking cDNA synthesis [1]. The majority of RNA-seq protocols require that RNA is reverse transcribed into cDNA before sequencing, but RT of tRNAs often results in truncated cDNA products that lack 5′ adapters and are never sequenced. Second, RT for RNA-seq is normally primed off of a ligated 3′ adapter, but mature tRNAs have compact secondary and tertiary structure with tightly base-paired 5′ and 3′ termini, which can prevent adapter ligation [2]. Hybridization methods such as microarrays and northern blot analysis have been used to quantify multiple tRNA species, but these approaches require prior knowledge of an organism's tRNA repertoire, and they may not able to evaluate expression levels of near-identical tRNAs because of cross-hybridization [3]. Additionally, RNA-seq has higher sensitivity for genes that are very highly or lowly expressed and can detect a much larger range of expression [4].

One significant technological advance in developing effective tRNA-seq protocols was the discovery that certain tRNA base modifications could be removed prior to RT using the demethylating enzyme AlkB from *Escherichia coli*. AlkB was shown to remove 1-methyladenine and 3-methylcytosine damage in single- and double-stranded DNA [5] and later utilized by Zheng et al. [6] and Cozen et al. [7] to remove some RT-inhibiting modifications present on tRNAs. In these studies, tRNA-seq libraries treated with AlkB had substantially higher abundance and diversity of tRNA reads. This was followed by the development of mutant forms of AlkB engineered to target specific modifications that appeared to be recalcitrant to demethylation by wild type AlkB treatment [6,8]. Although it is still not possible to remove all RT-inhibiting modifications with AlkB, the use of this enzyme represented a large step forward in tRNA sequencing.

In order to overcome inefficient adapter ligation, multiple methods have been developed involving modified adapter ligations [3,9-11] or a template-switching thermostable group II intron reverse transcriptase that eliminates the adapter ligation step entirely (DM-TGIRT) [6]. In the case of YAMAT-seq [3], Y-shaped DNA/RNA hybrid adapters are utilized to specifically bind to the unpaired discriminator base and the -CCA sequence motif added to the 3′ end of all mature tRNAs. The utility of adapter protocols that utilize CCA-complementarity and ligation of both the 5′ and 3′ adapters to intact RNAs is that full-length, mature tRNAs are preferentially sequenced even from total RNA samples. In

contrast, methods that ligate adapters in RNA/DNA step-wise fashion [10] or eliminate adapter ligation in the case of DM-TGIRT have been effective at broadly detecting tRNA gene transcription, but largely capture tRNA fragments. These truncated sequence reads are then difficult to confidently map to a single gene, requiring additional predictive models and bioinformatic solutions [12-14]. Despite the progress being made with AlkB to remove inhibitory modifications and methods such as YAMAT-seq to specifically target full-length tRNAs, no study has yet applied both approaches to combine the benefits of each.

Although post-transcriptional tRNA modifications have hindered efforts to sequence tRNAs, they are a universal and fundamental aspect of tRNA biology [15,16]. Base modifications are involved in tRNA folding and stability [17-19], translational accuracy and reading-frame maintenance [20,21], and tRNA fragment generation [16,22]. Not surprisingly given these fundamental roles, they have been implicated in numerous diseases [23], and there has been a long historical interest to identify, map, and quantify these modifications [24]. Traditional methods to identify base modifications involved tRNA purification and either direct sequencing and fingerprinting or complete digestion of RNAs into nucleosides followed by liquid chromatography-mass spectrometry (LC-MS) [25,26]. Soon after the discovery of reverse transcriptase enzymes, signatures of RT inhibition were also recognized as a means to map base modification positions [27]. Typical protocols involved primer extension assays and ^{32}P-labeled DNA primers to detect blocked or paused primer extension signals [28]. RT-based RNA modification detection assays were later extended through the use of chemical reagents that reacted with specific modified nucleotides to further confirm the identity of the modifications [1]. The advent of high-throughput sequencing has now facilitated the generation of entire maps or “indexes” of the modifications that cause RT-misincorporations across all tRNAs found in a species [29]. These signatures include fragment generation because of RT termination but also read-through misincorporations, i.e. base misincorporation and indels produced from the RT enzyme reading through a modified base [30,31]. Despite these advances, modification indexes have been produced for few species, and there has been limited investigation of tRNA modification profiles of isodecoders (tRNAs that share the same anticodon but have sequence variation at other parts along the tRNA). In addition, the function of some modifications is still unknown, and new modifications are still being discovered using a combination of RT-based modification mapping and mass spectrometry [32].

tRNA gene organization, expression, and modification patterns affect a wide diversity of biological processes, and the role of individual tRNAs has seen increasing interest in the past twenty years [33]. Species commonly have conserved multigene isoacceptor families (a group of tRNAs that are acylated with the same amino acid but can have different anticodons) [34], and there is growing evidence that certain tRNA genes have functionally unique roles [35,36]. One evolutionary lineage that has

particularly complex tRNA metabolism is angiosperms, or flowering plants. Plant nuclear tRNAs can be present in numerous copies, some of which are organized into large, tandemly repeated arrays [37,38]. Furthermore, the generation and function of tRNA-derived fragments (tRFs) has come under increasing scrutiny because their presence has been directly linked to plant cell growth and response to stresses [39-41]. Plants also have a high degree of tRNA compartmentalization and trafficking because of the presence of two endosymbiotically derived organelles, the plastid and the mitochondria. Each organelle has its own set of tRNAs, but plant mitochondria are exceptional compared to bilaterian animals in that they require extensive import of nuclear-encoded tRNAs into the mitochondrial matrix to maintain translation [42]. Additionally, plants have been shown to have some of the most post-transcriptionally modified tRNAs identified to date [43,44]. These modifications are increasingly being assigned new biological roles, making plants an ideal group to study the function of tRNA modifications.

Unfortunately, the reasons that make plant tRNA biology so fascinating (i.e. complex metabolism and extreme modification rates) have historically hindered investigation because of difficulties in applying high throughput sequencing and mapping methods, and recent advances in tRNA-seq such as the use of AlkB and CCA-complementary adapters have not yet been applied to plant systems. Here, we have combined these two tRNA-seq methods to detect full-length tRNAs, map them to unique reference gene sequences and identify RT-induced sequence variants in four different angiosperm species, focusing on the model system *Arabidopsis thaliana*.

RESULTS

YAMAT-seq captures mature, full-length tRNAs from the majority of *A. thaliana* tRNA reference genes

tRNAs were sequenced with a combination of two complementary methods, YAMAT-seq [3] and AlkB treatment [6,7] to generate full-length, mature tRNA reads from *A. thaliana* leaf total-cellular RNA. The demethylating enzyme AlkB was used in conjunction with YAMAT-seq's Y-shaped adapters that are complementary to the CCA motif found at the 3′ end of mature tRNAs to generate Illumina libraries composed almost entirely of tRNA sequences, with representatives from most genes in the reference *A. thaliana* database (all sequences in the database can be found in **supp. Table 1** with the corresponding tRNA gene identifiers). On average, 95% of all reads passed quality filters after trimming, and of those, greater than 99% of the reads mapped to a tRNA sequence when BLASTed to an *A. thaliana* tRNA database (e-value cutoff of 1e-6) (**supp. Table 2**). The remaining reads were largely derived from cytosolic/plastid rRNAs and adapter artifacts (**supp. Table 3**). Over 98% of the reads had a 3′-terminal

CCA sequence. Although the proportion of reads that matched the reference sequence length varied among isoacceptor families (**Figure 1**), the vast majority of the reads were 74-95 nt after adapter trimming (representing the expected range of mature tRNA sequence lengths in *A. thaliana*), indicating that our library size selection successfully targeted full-length tRNAs. A small fraction of the reads, even those with a CCA tail, were truncated. These fragments (defined as having less than 70% hit coverage to a reference tRNA) appear to be largely generated from a subset of tRNAs, particularly certain nuclear tRNA-Thr and tRNA-Ala genes. Many of these fragments appeared to involve sequences that inverted and re-primed off of short regions of sequence similarity on the complementary strand, producing a "U-turn" molecule.

AlkB treatment improves detection of *A. thaliana* reference tRNAs

Base modifications known to stall or terminate RT are prevalent in tRNAs [1]. The demethylating enzyme AlkB has been shown to effectively remove some of these modifications (N^1-methyladenosine [m^1A], N^3-methylcytosine [m^3C]), thereby making certain tRNAs more amenable to sequencing [6,7]. Our preliminary observations from testing AlkB treatments found a considerable effect on *A. thaliana* RNA integrity, with the majority of degradation occurring with exposure to the AlkB reaction buffer, regardless of whether the AlkB enzyme was included (**supp. Figure 1**). Even though this degradation predominantly affected larger transcripts and had little noticeable effect on the tRNA size fraction, we decided to use two types of negative controls in our experimental design to isolate the effects of AlkB from the buffer alone. Total RNA was either treated with the AlkB enzyme in the reaction buffer (AlkB+), treated with the reaction buffer alone (AlkB-), or left entirely untreated prior to RT. Libraries sequenced after the reaction buffer treatment alone and those that were entirely untreated did not substantially differ from each other in reference sequence detection and frequency (**supp. Figure 2**). However, treatment with AlkB resulted in better representation of the majority of cytosolic tRNAs and a moderate increase in the detection of some organellar tRNAs (**Figure 2, supp. Table 4**). Only 95-115 of the 183 nuclear tRNA reference genes were detected in untreated and AlkB- libraries (**supp. Table 5**), whereas AlkB+ treatment increased this range to 138-149. These reference coverage counts do not include tRNA reference genes that were exclusively detected with reads that were an equally good match to another reference sequence, which was the case for 4-13 genes per library (**supp. Table 5**). All 30 plastid and 17 of the 19 mitochondrial reference tRNAs were detected in at least one library with one mitochondrial tRNA-SerTGA gene only detected in a single AlkB+ library. No reads were detected for the mitochondrial genes tRNA-SerGCT-3360 and tRNA-SerGGA-3359. Ser and Tyr isoacceptors were the most likely to be undetected or have very low abundance in all libraries. Interestingly, three mitochondrial tRNA genes that are homologous to tRNA-PheGAA but have a mutated GTA (Tyr) anticodon and were previously annotated to be

pseudogenes in the *A. thaliana* mitochondrial genome (GenBank: NC_037304.1) were detected as mature, expressed tRNAs in this analysis, which is consistent with earlier detection of expressed copies of one of these genes [45]. In addition, the plastid-derived tRNA-Met in the mitochondrial genome was detected in this study whereas previous studies using hybridization methods alone failed to detect expression of this gene [46].

In addition to the wild type AlkB, an AlkB mutant (D135S) has been specifically engineered to remove the modification N^1-methylguanosine (m^1G), which is known to inhibit RT activity [6]. We performed tRNA-seq on three additional *A. thaliana* total-cellular RNA samples treated with either wild type AlkB, or a 2:1 ratio of wild type AlkB and D135S AlkB to test for further improvements in tRNA detection. We found only a moderate increase in the detection of a few genes when performing differential expression analysis on libraries treated with D135S (**supp. Table 6**). One D135S library did have a single read for a tRNA-TyrGTA gene that was undetected in the wild type libraries, but otherwise there was no increase in the number of genes detected in D135S libraries (**supp. Figure 3).**

tRNA-seq profiles are dominated by nuclear tRNA-Pro and plastid tRNA-GlyGCC genes in four angiosperm species

The number of reads mapped to each reference tRNA sequence varied drastically and was heavily skewed towards multiple nuclear tRNA-Pro genes and a plastid tRNA-GlyGCC **(Figure 3)**. Together, nuclear tRNAs-Pro and the plastid tRNA-GlyGCC sequences comprised 86-90% of all reads in AlkB+ libraries and 91-93% in untreated and AlkB- libraries. To test whether this dominance of tRNA-Pro and tRNA-GlyGCC reads was unique to *A. thaliana* or a more widespread pattern in flowering plants, tRNAs were sequenced from leaf total-cellular RNA from another rosid (*Medicago truncatula*), an asterid (*Solanum tuberosum*) and a monocot (*Oryza sativa*), using the same tRNA-seq method described above with wild type AlkB. The resulting reads from all three species showed a similarly extreme skew towards nuclear tRNA-Pro and plastid tRNA-GlyGCC (**Figure 4)**.

Persistent reverse transcription bias contributes to tRNA-seq coverage variation

The dominance of multiple nuclear tRNA-Pro genes and a plastid tRNA-GlyGCC gene was surprising because tRNA expression is generally expected to reflect codon usage after accounting for base-pairing modifications [47]. One possible artifactual source of variation in tRNA-seq read abundance is biased adapter ligation during library construction [48]. In order to determine the abundance of tRNA-derived cDNA molecules independent of adapter ligation, droplet digital PCR (ddPCR) was performed on reverse transcribed, unligated subsamples of the three original *A. thaliana* RNA replicates using internal primers for four tRNA genes (**supp. Table 7**). There was a strong correlation between counts per million tRNA-

seq reads (CPM) and ddPCR copies per nanogram (p = 0.03 and adjusted R^2 = 0.91; **Figure 5**). These data suggest that adapter ligation bias is not the primary determinant of tRNA-seq coverage skew as cDNA copy number in ddPCR is reflective of the number of final tRNA-seq Illumina reads.

An additional source of sequencing bias could result from preferential RT of tRNAs with fewer modifications or less inhibitory secondary structures. In order to quantify tRNA abundance in total-cellular RNA without the intermediate step of RT, northern blot analysis was performed by probing for four *A. thaliana* tRNAs representing a range of CPM values from the tRNA-seq data. All four labeled probes were hybridized to three RNA replicates as well as a dilution series of a complementary oligonucleotide in order to quantify the tRNA signal (**supp. Table 8**). The tRNA-Pro gene with the highest CPM showed the weakest hybridization signal, and the highly expressed plastid tRNA-GlyGCC did not have the strongest intensity of the two plastid tRNAs probed (**Figure 6)**. The concentration estimates from the northern blot analysis present a truly striking contrast with the tRNA-seq data because the read abundance for the sampled tRNA-ProTGG was approximately 70,000-fold higher than one of the other sampled nuclear tRNAs (tRNA-SerCGA). This massive incongruence strongly points to biased tRNA RT (even after AlkB treatment) as contributing to the extreme coverage variation found in our tRNA-seq data.

Because tRNAs are multicopy genes with high sequence similarity, a "mismatch oligo" with one or two noncomplementary nucleotides (relative to the probe) was included on each membrane to test for probe specificity and cross-hybridization. There was clearly a signal of cross-hybridization to these mismatch oligos (which was expected given the permissive hybridization temperature of 48°C). Thus, the signal for the RNA samples likely reflects some additional hybridization for other isodecoders with very similar sequence at the probe region. There are multiple isodecoders that are similar in sequence to tRNA-ProTGG-112 and tRNA-SerCGA-3245 at the probe region, and the mismatch oligos were designed to be identical to these similar isodecoders. There are no similar tRNAs to GlyGCC-3370 and IleGAT-3373 (the mismatch oligos for these genes were therefore "synthetic" with no biological match to another gene). Therefore, the already low hybridization signal for tRNA-ProTGG-112 and tRNA-SerCGA-3245 is likely an overestimation of expression, further exacerbating the incongruence between the northern blot analysis and the massive tRNA-seq CPM values for tRNA-ProTGG-112.

RT-induced misincorporations identify positions of base modifications in plant tRNAs

tRNA base modifications at the Watson-Crick face can interfere with the base pairing that is necessary for RT. Such modifications may not only stall or terminate reverse transcriptase activity but can also result in the misincorporation or deletion of nucleotides in the resulting cDNA. These misincorporations can be used to infer both the position and modification type, producing a modification map or "index" of a tRNA

[29,49,50]. Reads were globally aligned to reference sequences to identify all nucleotide positions that differed from the reference gene **(supp. Table 9**). Even after treatment with AlkB, a signal of RT misincorporation was still present in almost one-third of the tRNAs in at least one position. We identified a position as confidently modified if ≥30% of the mapped reads varied from the reference sequence at that position. There was evidence of multiple tRNAs being modified at the same site, with the same position being modified in up to 17% of all reference tRNAs (**Figure 7)**. Given that AlkB may act on only a subset of modifications on certain bases [29], it was unsurprising that positions with a modified T were largely insensitive to AlkB treatment (**Table 1**). Similar to work with RT-based modification detection in human cell lines and yeast [6,7,29], we found that demethylation treatment with AlkB had a strong effect at only certain tRNA positions (e.g., 9, 26, and 58). The most frequent type of misincorporation differed by both position and reference base, but Gs were most likely to be deleted whereas the other three bases were most likely to be misread as a substitution **(Table 2).**

In order to ensure that the majority of reads were being effectively mapped to a reference sequence, all mapped reads in the AlkB treated library 1 were BLASTed to the entire nuclear genome of *A. thaliana* to check if reads had a better hit to a nuclear location than one of the reference sequences in the database. Only 1106 reads in the AlkB1 treated library (0.097%) mapped to a genomic location that was not within 5bp window of a reference sequence. Of those, only 182 reads (0.016%) would have passed the 90% reference coverage threshold to be used for modification analysis (**supp. Table 10**). Given that we applied a 30% threshold to identify modified sites, mismapping of reads derived from unannotated nuclear tRNA genes or tRNA-like sequences appears to have a negligible effect on these predictions.

DISCUSSION

The promise and pitfalls of quantifying tRNA expression using tRNA-seq methods

In addition to the fundamental role of tRNAs as decoders of the genetic code, they are increasingly being recognized as integral players in a wide range of developmental, stress, tumorigenesis, biosynthetic, and amino acid delivery pathways [51-58]. As such, accurately detecting and quantifying tRNAs is key to gaining a more complete understanding of expression and regulation of numerous biological processes.

There have been substantial efforts in the past decade to make high-throughput sequencing methods more applicable to tRNAs. Nevertheless, we found that, even after treatment with AlkB and use of YAMAT adapters to specifically capture mature tRNAs, there still remained persistent RT-related sequencing bias, as evidenced by the extreme skew in tRNA-seq reads towards certain sequences (**Figures 3 and 4**). The observed dominance of multiple nuclear tRNA-Pro genes and a plastid tRNA-

GlyGCC likely does not reflect biological reality, as there was no agreement between the tRNA-seq read abundance estimates and the intensity of hybridization in northern blot analysis, where tRNAs are probed directly without an intervening RT step. Moreover, the massive inferred expression level of plastid tRNA-GlyGCC is at odds with previous 2D-gel analysis of purified chloroplast tRNAs, which did not identify tRNA-GlyGCC as unusually abundant [59-61]. “Jackpot” tRNAs with extremely and presumably artefactually high read abundance have been reported in prior tRNA-seq results [10,62], and the original YAMAT-seq paper reported a single tRNA-LysCTT gene having approximately 10-fold higher expression than the next most highly expressed gene [3].

We found that four divergent angiosperm species showed the same general tRNA-seq profile dominated by nuclear tRNA-Pro and plastid tRNA-GlyGCC, suggesting an underlying cause of sequencing bias that is broadly shared across angiosperms. One possibility is that the overrepresented genes are less modified than other tRNAs. The plastid tRNA-GlyGCC and almost all of the nuclear tRNA-Pro isoacceptors had no RT misincorporations (based on our 30% threshold) after AlkB treatment. However, many other tRNAs lacked a strong signal for modification at any position **(supp. Table 9)** but did not show the same high abundance as nuclear tRNA-Pro and plastid tRNA-GlyGCC. The fact that the original YAMAT-seq data generated from human cell lines shows a skewed distribution of entirely different isoacceptors suggests that tRNAs from different eukaryotic lineages vary in their properties that make them more (or less) amenable to sequencing.

Modifications that predominantly result in the termination of RT (“RT falloff”) without other signatures of RT inhibition (i.e. base misincorporations and indels) are not detectable by the YAMAT-seq method that we employed because molecules are not sequenced unless RT proceeds all the way through the tRNA and captures the 5′ adapter. Thus, variation across different tRNAs in the presence of modifications that induce RT falloff could be responsible for observed RT bias. Additionally, the effect of certain modifications on RT behavior has been shown to be dependent on the nucleotide 3′ of modifications in the template RNA [63], suggesting that even if a modification is detected and identified, it may have a different effect on expression analysis depending on the tRNA sequence. However, a study on cancer type tRFs found no association with the 5′ ends of tRFs and known modified positions [64], suggesting that the presence of base modifications was not leading to truncated reads. Thus, the RT of tRNAs may then be affected by secondary structure [1]. In the cases of the dominance of certain isoacceptor reads in this study, there may be otherwise ubiquitous secondary structure characteristics that are not present in plastid tRNA-GlyGCC or multiple nuclear tRNA-Pro isoacceptors, making these tRNAs more amenable to RT. These represent important areas of investigation to further understand and alleviate sources of bias in tRNA-seq methods.

The biases that we have identified make it clear that more work must be done in developing a tRNA sequencing method that can accurately quantify *absolute* levels of expression. Nevertheless, the combination of YAMAT-seq and AlkB treatment has a number of advantages and great promise for analyzing changes in *relative* tRNA expression across treatments, tissues or subcellular fractions. In particular, we were able to generate a high proportion of full-length tRNA reads which can be confidently mapped to a single reference gene. Mapping of tRNA-seq reads to loci can be problematic because of the large number of similar but non-identical tRNA gene sequences. Other tRNA sequencing methods that hydrolyze tRNA (Hydro-tRNA-seq) [13], or utilize a template-switching reverse transcriptase [DM-TGIRT-seq] [6] produce a large proportion of tRNA fragments that can ambiguously map to multiple genes. In general, it is common for tRNA-seq methods to report overrepresentation of certain isoacceptors [3,6,10,13,62], suggesting sequencing bias is a frequent problem. However, the skew we observed with YAMAT-seq was especially extreme (Figure 3). Therefore, alternatives to YAMAT-seq may benefit from a reduced level of sequencing bias, albeit often at the expense of only generating partial tRNA sequences. As there is increasing attention on the regulation and expression of specific tRNA genes [12,33], the combination of tRNA-seq methods that we employed could be effectively used for differential expression analysis when trying to tease apart the transcriptional activity and turnover of individual genes within large gene families.

It is important to note that plants, like many eukaryotes, have multiple copies of identical tRNA genes, and tRNA-seq cannot distinguish which copy, or copies, are expressed. This facet of tRNA expression is of growing interest in the regulation tRNA gene clusters such as the high-copy, tandemly repeated clusters of Ser, Tyr and Pro tRNA genes in *A. thaliana* [33,37]. The database of tRNA references used in this study represent collapsed identical sequences of tRNA genes (see **supp. Table 1** for all gene copy numbers by reference sequence). These identical copies can be found even across multiple genomes due to extensive organelle DNA insertions in the nucleus [65,66]. Although most intracellular gene transfers from organelles to the nuclear genome are thought to be nonfunctional, there is evidence in some eukaryotes that such organelle-derived sequences can be transcribed in the nucleus [67,68]. In the *A. thaliana* Col-0 accession, all but one identical, organellar-derived tRNA gene found in the nuclear genome are located in a large 620-kb insertion of mitochondrial DNA on chromosome two, comprising almost the entire mitochondrial genome [65]. This insertion is in a strongly heterochromatic, transcriptionally inactive region [69]. There are other, nonidentical of organelle-derived located in other nuclear locations, but no tRNA-seq reads from this analysis were found to be better matches to any of those nuclear genes than to the organelle copy (**supp. Table 10**).

Looking forward, technologies that directly analyze RNA molecules without the intermediate step of RT, such as nanopore sequencing [70], offer promise to accurately quantify tRNA expression without

the confounding effects of RT bias. However, nanopore technologies still require substantial development to achieve sufficient sequencing accuracy and differentiation of tRNA species, especially in the context of the extensive base modifications present in tRNAs [9]. In the interim, our combined approach may represent one of the more effective means to quantify relative tRNA expression changes at a single-gene level.

The landscape of base modifications in plant tRNAs

Although base modifications likely contribute to biased quantification of tRNA expression, there remains a large benefit of utilizing a RT-based tRNA-seq methods because RT misincorporation behavior provides insight into the location and identity of base modifications. In addition to modifications that terminate cDNA synthesis, RTs can read through methylations at the Watson-Crick face with varying efficiencies, resulting in misincorporations and indels in the cDNA [50,71,72]. Here, we took advantage of RT-based expression analysis to present one of the most extensive modification landscapes of angiosperm tRNAs to date. The modification peaks produced from sequencing complete *A. thaliana* tRNAs (**Figure 7**) provide an informative comparison to the annotated modification indexes previously generated for human tRNAs [29] as well as in the recently published PRMdb database [73] of predicted plant tRNA base modifications. Given the widely conserved presence of some tRNA modifications known to inhibit RT [74], it was unsurprising that the modification indexes from plants and humans shared many similarities, including a high rate of modification at nucleotide positions 9, 20, 26, 34, 37, and 58 (**Figure 8**). We reported modifications based on their actual position in tRNA genes. Thus, in some cases, the same modification at functionally analogous sites in different tRNAs can be represented by peaks spanning two or three nt positions due to differences in tRNA length.

Multiple software programs have been developed to predict RNA base modifications using misincorporation signatures in transcripts. For example, HAMR ([49]) has been used by the database PRMdb[73] to predict modifications in plant RNAs. Many positions are predicted to be modified by both HAMR and our analysis, including positions 9, 20, 26, 37, and 58 (**supp. Figure 4).** There do, however, exist some differences between the two methods. The modification index generated in this analysis has strong misincorporation signals in the variable loop (positions 46-48), which are missing from the PRMdb analysis. This may be partly due to the PRMdb tRNA database containing predominantly nuclear tRNAs, whereas our analysis used all nuclear, plastid and mitochondrial tRNAs to construct a modification index. In particular, plastid and mitochondrial tRNAs appear to be very commonly modified at positions 46-48 (**supp. Figure 5**). Even so, our analysis found multiple nuclear tRNA-Thr, tRNA-Tyr and tRNA-Asn isodecoders with misincorporation signals at positions 46 and 48 that were not reported by HAMR.

Similarly, our dataset lacks modification predictions at positions 67-69, likely because of low coverage of some tRNA-Ser and tRNA-Leu isodecoders.

Previous work [6] has confirmed position 9 as frequently having an m^1G modification and position 58 having an m^1A modification. These modifications are known targets of AlkB demethylation, and we found that AlkB treatment resulted in almost complete reduction of the modification index peaks at positions 9 and 58 (some of this signal appears at position 57 in our modification index because of differences in tRNA length) (**Figure 7**), adding support for the existence of these same modifications in plants. Two consecutive dihydrouridines are commonly found in the D loop around position 20 in eukaryotes [75], a modification motif that may also be common in plant tRNAs, as we found a modification peak around positions 19-21. Position 26 is often modified to N2,N2-dimethylguanosine ($m^2{}_2G$), which also appears to be mildly sensitive to AlkB demethylation treatment [29]. We found a modification peak at position 26 and a small reduction in RT misincorporation rate after AlkB treatment in *A. thaliana*. Adenosine-to-inosine modifications are frequently found at position 34 in eukaryotic tRNAs [76]. RT of inosine is expected to cause A-to-G changes because inosine base pairs with C instead of T [1]. Indeed, we found almost 100% misincorporation of G when the reference was A at position 34. It is important to note that some tRNA modifications may not be present in all copies [77] or may not always cause RT misincorporation [50]. Thus, our modification index with a cutoff of ≥30% misincorporation is focusing on highly modified positions in the corresponding tRNAs [29].

Our analysis also identified clear differences in the modification index between plants vs. humans. In particular, we found a strong misincorporation peak at positions 46-48 for T reference nucleotides that was not observed in humans (**Figure 8**) [29]. The misread Ts at position 46-48, could correspond to dihydrouridines, which are commonly found in the variable loop (position 47 in standard tRNA nomenclature) of eukaryotic and plastid tRNAs [78,79]. Taken together, over 14% of all *A. thaliana* reference tRNAs appear to have a modified T at the position, including all cytosolic and plastid tRNA-Ile isoacceptors, as well as all cytosolic tRNA-ThrAGT, tRNA-AsnGTT, and tRNA-TyrGTA isodecoders. This trend of frequent modification at position 47 was conserved in all angiosperm species tested (**supp. Figure 6, supp. Table 11**). It is interesting to note that the misincorporation profile (deletions and T-to-C misincorporations) for dihydrouridines at positions 19-21 is similar to the misincorporation profiles that we observe for positions 46-48. Many *A. thaliana* tRNAs have a G in their reference sequence at position 46 and frequently exhibited a deletion at this site. This pattern likely reflects a modification to m^7G, which is predicted to be common in cytosolic Viridiplantae tRNAs at that position [43]. Adding to this prediction is the complete lack of an AlkB effect on the modification peak at position 46, as AlkB is not expected to remove m^7G modifications. Our analysis demonstrates how tRNA-seq methods can be used to elucidate similarities and differences in global tRNA modification

patterns between divergent taxonomic groups. These approaches could be extended to assess whether different tissues or even subcellular compartments depend more heavily on certain tRNA modifications for functionality.

In addition to generating a general modification index across the entire tRNA population, this tRNA-seq method can be used to identify the modification profiles of individual tRNAs or specific groups of tRNAs. In other eukaryotes, organellar tRNAs have been found to be less modified than their cytosolic counterparts [42,43]. In agreement with this, we detected fewer modifications on average in *A. thaliana* organellar tRNAs. Cytosolic tRNAs with a misincorporation signal had an average of 3.0 modifications in AlkB- libraries, whereas organelle tRNAs with misincorporations had an average of 2.1 modifications. Likely owing to greater rates of modification in cytosolic tRNAs, AlkB treatment had a substantially larger effect on improving the detection of cytosolic tRNAs than on organellar tRNAs (**Figure 2**).

It is sometimes assumed that tRNAs with the same anticodon but different body sequences will have similar modification patterns [12]. Nevertheless, we found examples of large differences in modification patterns among isodecoders. For example, the large nuclear tRNA-GlyGCC isodecoder family has a range of one to six inferred modifications depending on the reference sequence and show a range of sensitivity to AlkB treatment. That some tRNAs, even those with the same anticodon, have such radically different modification patterns represents an interesting and largely unexplored facet of tRNA biology, and offers support to the finding that isodecoders may differ in function [80].

We found that AlkB treatment sometimes created novel RT misincorporations instead of restoring the correct nucleotide. For example, the C in the anticodon of the plastid tRNA-IleCAT is known to be modified to lysidine, which has base-paring behavior like that of U [81,82], and we found this modification to be consistently misread as a deletion in AlkB+ libraries, whereas it was detected as a T, or even rarely as a C, in untreated libraries. Additionally, an increase in abundance of some tRNA fragments was found in AlkB+ libraries (**supp. Table 12**) and spatially associated with modified bases. For example, the "U-turn" fragments generated from tRNA-ThrCGT-3316 invert at what appears to be a modified G25, and the abundance of these fragments increased with AlkB treatment. In addition, treatment with AlkB as well as the AlkB reaction buffer had a significant effect on RNA integrity (**supp. Figure 1**), meriting further investigation into the cause of RNA integrity loss and whether this degradation is system specific.

Historically, comprehensive mapping of base modifications has been laborious and depended on tRNA purification followed by digestion and mass spectrometry [83]. Quickly and accurately identifying tRNA modifications through RT-based methods makes it possible to test hypotheses regarding the presence and function of modifications at phylogenetic, developmental, and subcellular levels. tRNAs are

by far the most post-transcriptionally modified gene class [84]. Research is just now starting to tease apart the many roles these modifications play [18,85-87]. Not all modifications are detectable through sequencing, but it has been estimated that RT-based methods can identify approximately 25% of tRNA modifications in humans [29], and the modifications that can be detected by sequencing are known to have critical roles in tRNA folding, rigidity and stability [71]. Given the diversity of tRNA modifications, the full scope of RT behavior and the effect of AlkB treatment on tRNA base modifications has yet to be fully explored, but work is currently being done to further characterize RT behavior when encountering specific modifications [50,88,89]. Moving beyond RT-based methods to direct sequencing of tRNAs and their modifications is another exciting development on the horizon. However, major advances must still be made in techniques such as nanopore sequencing because current technologies cannot be used to reliably sequence tRNA molecules in biological total RNA samples because they are likely unable to discriminate between more than two tRNAs at a time [90,91]. Additionally, efforts to characterize the resistance properties of alternative tRNA base modifications in nanopores are underway [92] but still in their infancy. Thus, along with the promise of additional technologies on the horizon, current RT-based methods offer easily accessible and exciting tools to study base modification profiles of full-length tRNA sequences and how they vary across taxa, treatments, tissues, and subcellular locations.

METHODS

Plant material and growth conditions

For tRNA-seq material, *Arabidopsis thaliana* Columbia ecotype (Col-0) was grown in Pro-Mix BX General Purpose soil supplemented with vermiculite and perlite. Plants were germinated and kept in growth chambers at Colorado State University 23°C with a 16-hr/8-hr light/dark cycle (light intensity of 100 $\mu E \cdot m^{-2} \cdot s^{-1}$). *Solanum tuberosum* var. Gladstone (PI: 182477), *Oryza sativa* var. Ai-Chiao-Hong (PI:584576) and *Medicago truncatula* (PI:660383) were acquired from the U.S. National Plant Germplasm System (https://www.ars-grin.gov/npgs/). Tissue culture plantlets of *S. tuberosum* were transferred to Pro-Mix BX General Purpose soil supplemented with vermiculite and perlite and kept in growth chambers with the same settings as above. Seeds of *O. sativa,* and *M. truncatula* were germinated in SC7 Cone-tainers (Stuewe and Sons) on a mist bench under supplemental lighting (16-hr/8-hr light/dark cycle) in the Colorado State University greenhouse, then moved to a growth chamber with the same settings as above after 4 weeks. For northern blot analysis, *A. thaliana* Col-0 was grown in a 12-hr/12-hr light/dark cycle in growth chambers at the Institute of Molecular Biology of Plants.

RNA extraction

For our primary tRNA-seq experiment and ddPCR analysis, RNA from *A. thaliana* and *O. sativa* was extracted from 7-week-old leaf tissue. RNA from *M. truncatula* was extracted from 3-week-old leaf tissue, and RNA from *S. tuberosum* plantlets was extracted from leaf tissue 3 weeks after transfer into soil. Extractions were performed using a modified version of the Jordon-Thaden et al. [93] protocol. In brief, tissue was frozen in liquid nitrogen and pulverized with a mortar and pestle and then vortexed and centrifuged with 900 µl of hexadecyltrimethylammonium bromide (CTAB) lysis buffer supplemented with 1% polyvinylpyrrolidone (PVP) and 0.2% β-mercaptoethanol (BME). Samples were then centrifuged, and the aqueous solution was removed. 800 µL of a chloroform:isoamyl alcohol (24:1) solution was added, mixed by inverting and centrifuged. The aqueous phase was removed, and 900 µl of TRIzol was added. The solution was then mixed by inversion and centrifuged. The aqueous phase was removed, 200 µl of chloroform was added, and the solution was then mixed and centrifuged again. The aqueous phase was removed followed by isopropanol RNA precipitation, and cleaned pellets were resuspended in dH_2O. RNA was checked for integrity with a TapeStation 2200 and purity on a Nanodrop 2000. Later RNA extractions from *A. thaliana* for the AlkB D135S tRNA-seq experiment (8-week-old tissue) and northern blot analysis (4-week-old tissue) were performed with a simplified protocol. Leaf tissue was frozen in liquid nitrogen and pulverized with a mortar and pestle prior to the addition of TRIzol but otherwise followed the TRIzol manufacturer's RNA extraction protocol. Three biological replicates (different plants) were used for the *A. thaliana* experiments, and a single sample was used for each of the three other angiosperms.

AlkB purification

Plasmids containing cloned wild type AlkB protein (pET24a-AlkB deltaN11 [plasmid #73622]) and D135S mutant protein (pET30a-AlkB-D135S [plasmid #79051]) were obtained from Addgene (http://www.addgene.org/). Protein was expressed and purified at CSU Biochemistry and Molecular Biology Protein Expression and Purification Facility. Cells were grown at 37°C to an OD_{600} of 0.6, at which time, 1 mM of isopropyl-β-D-thiogalactoside was added, and temperature was lowered to 30°C. Cells were harvested after 3 hr by centrifugation and resuspended in 10 mM Tris (pH 7.3), 2 mM $CaCl_2$, 300 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, and 1 mM BME. Resuspension was homogenized by sonication, and lysate was recovered by centrifugation. The supernatant was loaded onto HisTrap HP 5 ml columns (GE Healthcare) and was washed and eluted by a linear gradient of 0-500 mM imidazole. The fractions containing AlkB were pooled and concentrated with ultrafiltration using Amicon Ultra-15 MWCO 10kDa (Millipore). The concentrated sample was then loaded onto HiLoad 16/60 Superdex 200 prep grade size exclusion column (GE Healthcare) in 20 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT,

and 10% glycerol. The fractions containing AlkB were pooled and concentrated, aliquoted, flash-frozen in the presence of 20% glycerol and stored at -70°C.

Demethylation reaction

AlkB reactions were performed using a modified version of existing protocols [6,7,94]. Demethylation was performed by treating 10 µg of total cellular RNA with 400 pmols of AlkB in a reaction volume of 80 µl containing: 70 µM ammonium iron(II) sulfate hexahydrate (Sigma-Aldrich, 203505), 0.93 mM α-ketoglutaric acid disodium salt dihydrate (Sigma-Aldrich, 75892), 1.86 mM ascorbic acid (EMD, AX1775-3), and 46.5 mM HEPES (pH 8.0-HCl, Corning, 61-034-RO), incubated at 37°C for 60 min in 0.2 ml PCR strip tubes (Fisherbrand) in a preheated thermal cycler (Bio-Rad, C1000 Touch). The reaction was quenched by adding 4 µl of 100 mM EDTA (Invitrogen, AM9262) followed by a phenol-chloroform RNA extraction, ethanol precipitation with the addition of 0.08 µg of RNase-free glycogen (Thermo Scientific, R0551), and resuspension in RNase-free water (Invitrogen, 10-977-023). Prior to the reaction, the AlkB enzyme was diluted to 100 pmol/µl with a dilution buffer containing 10 mM Tris-HCl pH 8.0 (National Diagnostics, EC-406), 100 mM NaCl (Fisher, S271-3), and 1 mM BME (Acros Organics, 125472500). The reaction buffer was prepared fresh with reagents added in the order listed above. The diluted AlkB enzyme was added to the reaction buffer followed by the RNA, and the reaction was brought to the final volume using RNase-free dH_2O. RNA integrity was checked on a TapeStation 2200. The same procedure was followed in parallel for AlkB- control libraries except that the AlkB enzyme in the reaction volume was replaced with dH_2O.

Illumina tRNA-seq library construction

All adapter and primer sequences used in library construction can be found in **supp. Table 13**. In order to remove amino acids from the mature tRNAs (deacylation), demethylated or control RNA was incubated in 20 mM Tris HCl (pH 9.0) at 37°C for 40 min. Following deacylation, adapter ligation was performed using a modified protocol from [3]. A 9 µl reaction volume containing 1 µg of deacylated RNA and 1 pmol of each Y-5′ adapter (4 pmols total) and 4 pmols of the Y-3′ adapter was incubated at 90°C for 2 min. 1 µl of an annealing buffer containing 500 mM Tris-HCl (pH 8.0) and 100 mM $MgCl_2$ was added to the reaction mixture and incubated for 15 min at 37°C. Ligation was performed by adding 1 unit of T4 RNA Ligase 2 enzyme (New England Biolabs) in 10 µl of 1X reaction buffer and incubating the reaction at 37°C for 60 min, followed by overnight incubation at 4°C. We found that adapter ligation with the input of 80 pmols of adapters called for in the original protocol resulted in an excess of adapter dimers and other adapter-related products after PCR and that using $1/10^{th}$ of that adapter concentration

maximized yield in the expected size range for ligated tRNA products, while minimizing adapter-related products.

RT of ligated RNA was performed using SuperScript IV (Invitrogen) according to the manufacturer's protocol. Briefly, 1 µl of 2 µM RT primer, and 1 µl of 10 mM dNTP mix was added to 11 µl of the deacylated RNA from each sample. The mixture was briefly vortexed, centrifuged and incubated at 65°C for 5 min. Then, 4 µl of 5X SSIV buffer, 1 µl 100 mM DTT, 1 µl RNaseOUT, and 1 µl of SuperScriptIV were added to each reaction. The mixture was then incubated for 10 min at 55°C for RT and inactivated by incubating at 80°C for 10 min.

The resulting cDNA was then amplified by polymerase chain reaction (PCR) in a 50 µl reaction containing 7 µl of the RT reaction, 25 µl of the NEBNext 2X PCR Master Mix, 1 µl of the PCR forward primer, 1 µl of the PCR reverse primer, and 15.5 µl dH_2O. Ten cycles of PCR were performed on a Bio-Rad C1000 Touch thermal cycler with an initial 1 min incubation at 98°C and 10 cycles of 30 s at 98°C, 30 s at 60°C and 30 s at 72°C, followed by 5 min at 72°C.

Size selection of the resulting PCR products was done on a BluePippin (Sage Science) with Q3 3% agarose gel cassettes following the manufacturer's protocol. The size selection parameters were set to a range of 180-231 bp, with a target of 206 bp. Size-selected products were then cleaned using solid phase reversable immobilization beads and resuspended in 10 mM Tris (pH 8.0).

Sequencing and read processing

The nine original *A. thaliana* tRNA-seq libraries (three AlkB+, three AlkB-, and three entirely untreated) were single-indexed and sequenced on an Illumina MiSeq with single-end, 150 bp reads. Libraries from *M. truncatula*, *O. sativa, S. tuberosum*, and the *A. thaliana* D135S AlkB mutant and wild type AlkB libraries were dual-indexed and sequenced on an Illumina NovaSeq 6000 with paired-end, 150-base pair reads. Sequencing reads are available via the NCBI Sequence Read Archive under BioProject PRJNA562543. Adapters were trimmed using Cutadapt version 1.16 [95] with a quality-cutoff parameter of 10 for the 3′ end of each read. A minimum length filter of 5 bp was applied to reads from the MiSeq sequencing platform. Read length filters of a minimum of 50 bp and a maximum of 95 bp were applied to reads produced from the NovaSeq 6000, as a much larger percentage of the reads from those libraries were <20 bp after adapter trimming. For paired-end data, BBMerge from the BBTools software package was used to merge R1 and R2 read pairs into a consensus sequence [96]. Identical reads were summed and collapsed into read families using the FASTQ/A Collapser tool from the FASTX-Toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html).

Sequence alignments and mapping

Mapping for all processed and collapsed reads was performed with a custom Perl script using reference tRNA databases from the PlantRNA website (http://seve.ibmp.unistra.fr/plantrna/, downloaded 9/8/2018). The database for *A. thaliana* comprises 579 nuclear tRNA genes, 23 mitochondrial tRNAs, and 37 plastid tRNAs that collapse down to 232 unique read families. This database is manually curated and updated to reflect current tRNA biology in model plant species. Each collapsed read family was BLASTed (blastn, 1e-6) against a complete set of (non-identical) nuclear, mitochondrial and plastid reference tRNA sequences from the corresponding species, and the best match was assigned to each read family. The read count of all assigned read families was then summed for each reference sequence. When a read family was an equally good match to multiple references, the read count was divided by the total number of tied references. Multiple BLAST statistics were recorded for each read family/reference hit pair including e-value, hit score, hit length, and percent identity. The read family sequences, tRNA identifier and the gene copy number can be found for every database sequence in **supp. Table 1.**

Because we were mainly interested in mature tRNA sequences, only reads that had 80% or greater hit coverage to a reference sequence were used in downstream analysis. Complete datasets including reads that fell below the 80% hit coverage threshold were used to determine read length distributions and the proportion of reads that did not BLAST to a tRNA reference sequence.

Efficiency of mapping was calculated by BLASTing the first AlkB treated library (AlkB1) to all *A. thaliana* chromosomes and custom Perl scripts were used to check for reads that had a better hit (raw bit score) to any genomic location than the mapped reference sequence. The coordinates of the genomic positions of the hit of any reads with a better hit were then compared to the genomic positions of the reference sequences using a custom Perl script.

Generation of modification index

In order to capture all possible misincorporations and indels including those that could have occurred at the end of the reads, a custom Perl script and the alignment software MAFFT version 7.407 [97] was used to produce full-length alignments of reads to their top BLAST reference tRNA hit and identify all base substitutions and indels in the resulting alignment. Only reads that BLASTed to their reference sequence with an e-value of 1e-10 or less and had 90% hit coverage to the reference were retained for modification index calling. Only positions that were covered by greater than five reads were used for modification index calling, and a position was considered confidently modified if ≥30% of the mapped reads at that position differed from the reference nucleotide by either a substation or deletion. Scripts used for read processing, mapping, and sequence analysis are available via GitHub (https://github.com/warrenjessica/YAMAT-scripts).

The modification index of predicted tRNA modifications from PRMdb [73] was generated by downloading all tRNA modification data for *Arabidopsis thaliana* found at http://www.biosequencing.cn/PRMdb/index.html (downloaded on 5/5/2020). The total number of tRNAs with a modification at each position was summed and divided by the total number of tRNAs in that database.

ddPCR

To generate cDNA for ddPCR quantification, 1 µg of RNA from each of the three original AlkB+ *A. thaliana* samples was treated with DNase I (Invitrogen) according to the manufacturer's protocol. For each sample, 146 ng of DNase-treated RNA was then subjected to RT using iScript Reverse Transcription Supermix (BioRad) according to manufacturer's protocol in a 20-µl reaction volume. cDNA was then diluted in a four step concentration series to 0.001 ng/µl, 0.01 ng/µl, 0.1 ng/µl, and 1 ng/µl. ddPCR was performed with each of the four primer pairs (**supp. Table 6**) using this concentration series of template cDNA to determine ideal concentration. All ddPCR amplifications were set up in 20-µL volumes with Bio-Rad QX200 ddPCR EvaGreen Supermix and a 2 µM concentration of each primer before mixing into an oil emulsion with a Bio-Rad QX200 Droplet Generator. The final template input amounts were 0.1 ng for amplification of tRNA-GlyGCC-3370 and tRNA-CysGCA-3354, 0.001 ng for amplification of tRNA-ProTGG-112, and 1 ng for amplification of tRNA-GlyCCC-93. Amplification was performed on a Bio-Rad C1000 Touch Thermal Cycler with an initial 5 min incubation at 95°C and 40 cycles of 30 s at 95°C and 1 min at 60°C, followed by signal stabilization via 5 min at 4°C and 5 min at 95°C. The resulting droplets were read on a Bio-Rad QX200 Droplet Reader. Copy numbers for each PCR target were calculated based on a Poisson distribution using the Bio-Rad QuantaSoft package.

Northern blots

RNAs and synthesized oligonucleotides (**supp. Table 7**) were separated on 15% (w/v) polyacrylamide gels. Gels were then electrotransfered onto Hybond-N^+ nylon membranes (Amersham) for 90 min at 300mA/250V and UV-crosslinked. All membranes were hybridized with ^{32}P-radiolabeled oligonucleotide probes (**supp. Table 7**) at 48°C in a 6X saline-sodium citrate (SSC) buffer with 0.5% sodium dodecyl sulphate (SDS) solution for 14 hr. Hybridization was followed by two washes (10 min) with 2X SSC buffer at 48°C buffer followed by one wash (30 min) at 48°C in 2X SSC with 0.1% SDS. Membranes were imaged on a Typhoon FLA 9500 biomolecular imager (GE Healthcare Life Sciences) after 16 hr of exposure on film.

Northern blots were analyzed using ImageJ, version Java 1.8.0_172 (64-bit) (https://imagej.nih.gov/ij/) to quantify signal intensity for each band. Signal peaks were quantified by first

drawing a straight (0.00 degree) line beginning at the farthest left point of contact of the signal to the opposite side of the window. To define only the major signal peak of interest, two lines were drawn perpendicular to the horizontal. The position of these lines was determined by eye based on where the peak's curves began to smooth. The coordinates of these vertical boundaries were used for all subsequent plots of the same membrane. The area was then calculated between the horizontal line and the peak signal bounded by the two vertical lines. Using the known concentration of the loaded oligos and the corresponding signal area, the estimated pmols of each tRNA target was calculated by fitting a standard curve to the $\log_{10}$ values of signal intensity and oligo concentration (**supp. Table 7**).

Statistics and figure generation

Differential expression analysis between the AlkB treated and untreated libraries in the original analysis as well as the wild type and D135S mutant AlkB libraries was done with the R package EdgeR version 3.24.3 [98], using a dataset with only reads that had 100% hit coverage to a reference sequence. A linear regression was performed in R with the lm function to test the relationship between tRNA-seq CPM values and ddPCR copy-number estimates per ng of input RNA.

ACKNOWLEDGEMENTS

We thank Deyu Li for providing valuable guidance in optimizing the AlkB reaction and all of the research personnel at the Institute of Molecular Biology of Plants for graciously hosting J.M.W. We would also like to thank Dr. Carol Wilusz and Adam Heck for test RNA samples and valuable discussions, and Dr. John Hunt and Dr. Victor Wong at Columbia University for sending samples of AlkB for initial testing. Two anonymous reviewers provided helpful comments on an earlier version of this manuscript. This work was supported by a Chateaubriand Fellowship from the Embassy of France in the United States, an NSF GAUSSI graduate research fellowship (DGE-1450032), NSF award IOS-1829176, and start-up funds from Colorado State University.

DATA AVAILABILITY STATEMENT

Sequencing reads are available via the NCBI Sequence Read Archive under BioProject PRJNA562543.

DISCLOSURE STATEMENT

The authors declare no financial conflict of interest.

Table 1| Total proportion of reference nucleotides misread ≥30% of the time in *Arabidopsis thaliana* libraries.

Library	Reference			
	A	C	G	T
AlkB+ (rep 1)	0.0187	0.0028	0.0147	0.0149
AlkB+ (rep 2)	0.0168	0.0028	0.0113	0.0140
AlkB+ (rep 3)	0.0136	0.0011	0.0100	0.0112
AlkB- (rep 1)	0.0350	0.0038	0.0241	0.0100
AlkB- (rep 2)	0.0361	0.0032	0.0258	0.0089
AlkB- (rep 3)	0.0299	0.0036	0.0224	0.0098
Untreated (rep 1)	0.0347	0.0038	0.0237	0.0112
Untreated (rep 2)	0.0294	0.0032	0.0213	0.0086
Untreated (rep 3)	0.0339	0.0041	0.0213	0.0100

Table 2| Total misincorporation frequency of tRNA reference bases across all nine *Arabidopsis thaliana* tRNA-seq libraries (see Table 1).

Misincorporation	Reference A	Reference C	Reference G	Reference T
A	**-**	**0.141**	**0.077**	**0.037**
C	**0.041**	**-**	**0.075**	**0.275**
G	**0.294**	**0.15**	**-**	**0.315**
T	**0.535**	**0.508**	**0.209**	**-**
Deletion	**0.129**	**0.202**	**0.639**	**0.373**

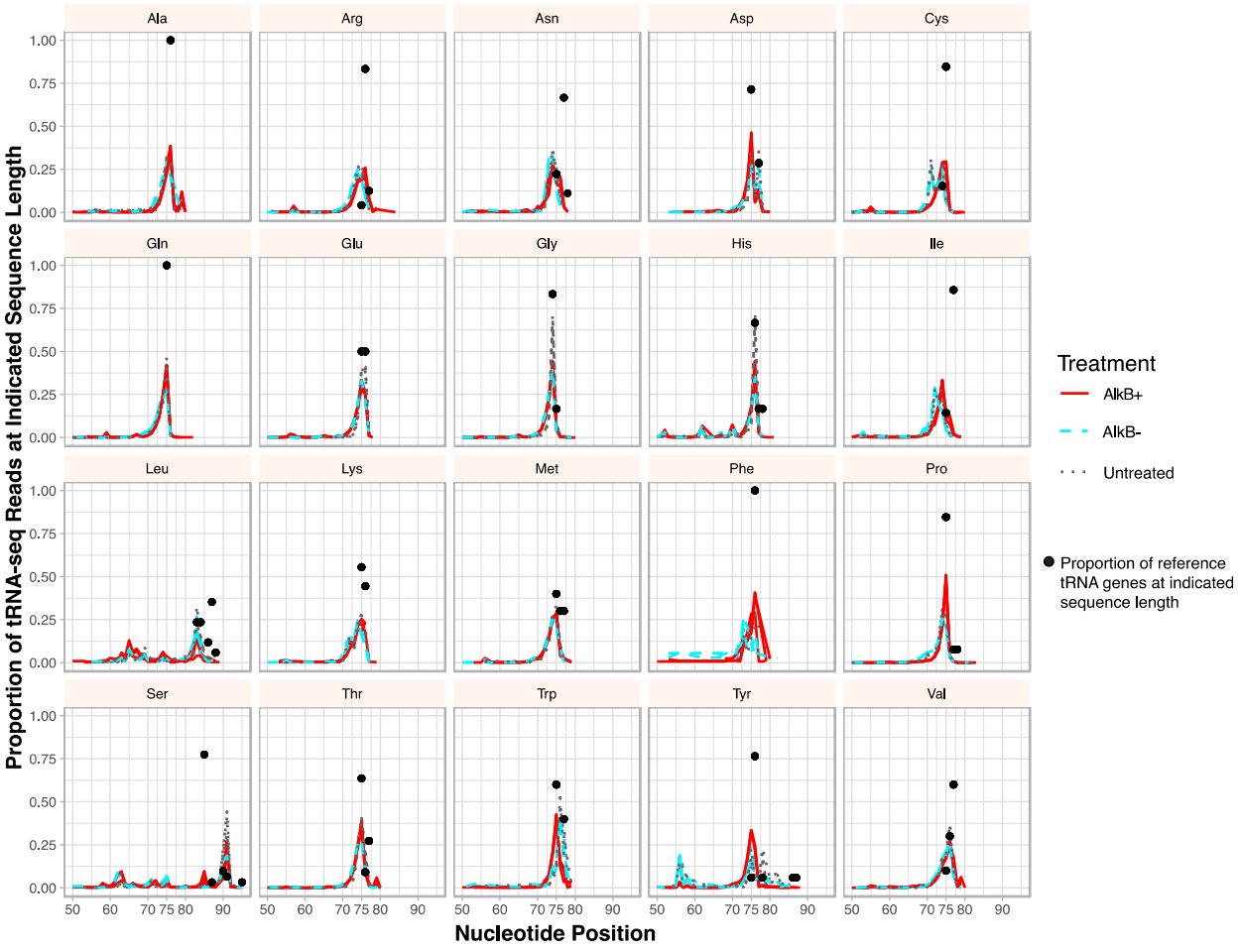

Fig. 1| The proportion of tRNA-seq reads at different lengths based on isoacceptor family. Black dots represent the proportion of isoacceptor reference sequences at indicated length (e.g. 100% of the Ala-tRNA reference genes are 76 nt in length). Three biological replicates are plotted for each treatment (AlkB+, AlkB- and untreated). Note that the dominant peak for read length often corresponds to the length of the reference sequence(s).

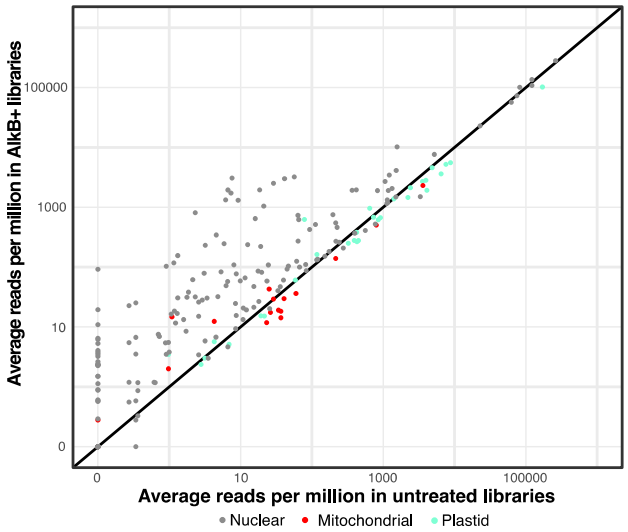

Fig. 2| Demethylation treatment with AlkB increases the proportion of tRNA-seq reads for many tRNAs (points falling above the 1:1 line). Average read counts per million across three biological replicates are shown for each unique *A. thaliana* tRNA reference sequence in AlkB+ vs. untreated libraries.

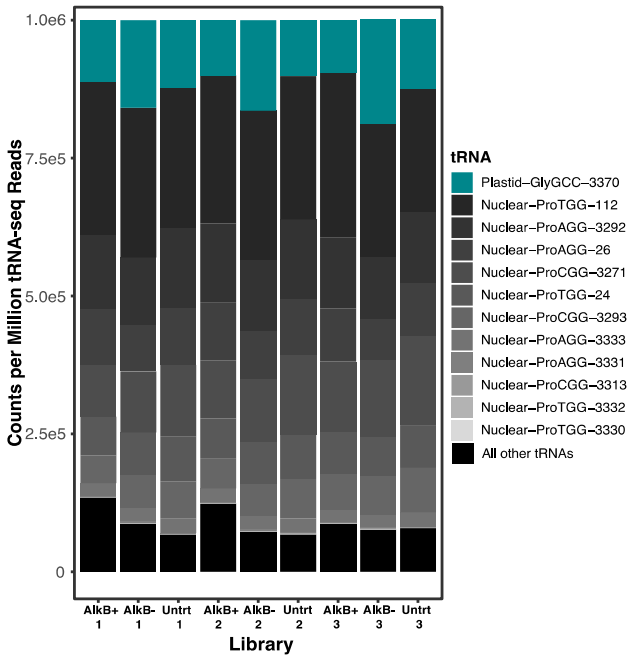

Fig. 3| tRNA-seq read abundance in *A. thaliana* tRNA is dominated by nuclear tRNA-Pro and plastid tRNA-GlyGCC genes. All nuclear tRNA-Pro genes are shown and indicated by gray colors, with the darkest grays indicating the highest abundance, and the plastid tRNA-GlyGCC is indicated by teal. All other tRNAs have been grouped and shown in black.

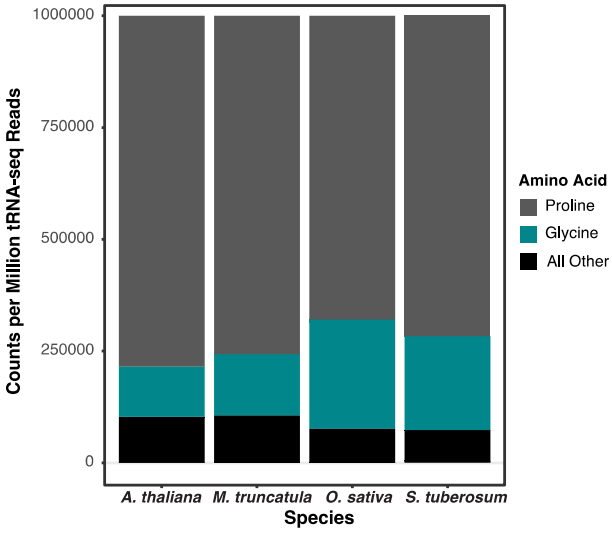

Fig. 4| tRNA-seq read abundance is similarly dominated by only two isoacceptor families (tRNA-Pro and tRNA-Gly) in four angiosperm species.

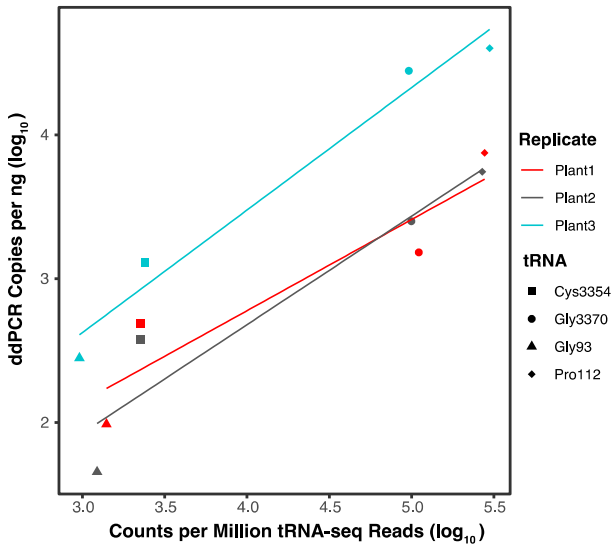

Fig.5| Droplet digital PCR (ddPCR) copies correlate with number of tRNA-seq reads. ddPCR copies per ng of cDNA plotted against counts per million tRNA-seq reads for four *A. thaliana* tRNA genes. Adjusted R^2 values for separate linear regressions on biological replicates 1, 2, and 3 were 0.79, 0.85, and 0.96, respectively. When data points were averaged across biological replicates, linear regression yielded an adjusted R^2 of 0.91 and a *p*-value of 0.03.

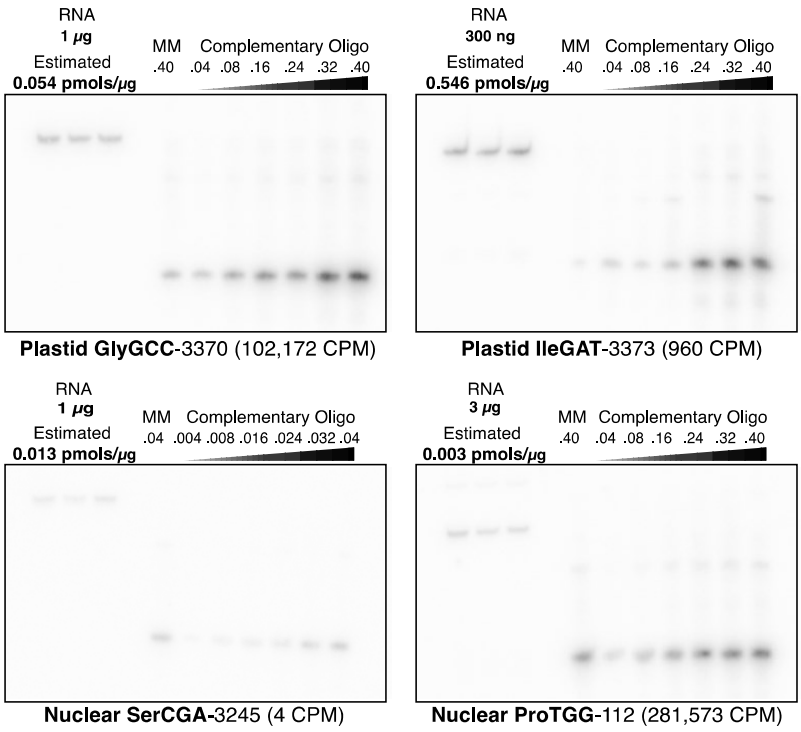

Fig. 6| Northern blot analysis does not show the same expression dominance of nuclear tRNA-Pro and plastid tRNA-GlyGCC that was observed with tRNA-seq. Four different *A. thaliana* tRNA genes were probed from total cellular RNA and quantified. Each tRNA target membrane had three replicates of total *A. thaliana* RNA, which was quantified using a dilution series of a synthesized 38-nt oligonucleotide complementary to the corresponding probe. A single lane of a 38-nt mismatched oligonucleotide (MM), which had one or two non-complementary nucleotides relative to the probe, was also included on each membrane to test for cross hybridization of probes. The amounts of total RNA and oligonucleotides loaded on each blot were varied according to expected hybridization signal based on preliminary analyses. The mass of RNA and pmols of oligonucleotides are indicated above each corresponding lane. The estimated concentration of the target tRNA based on analysis of signal intensity in ImageJ is reported as pmols per μg of input RNA. The average Illumina tRNA-seq read counts per million (CPM) is indicated parenthetically for each tRNA.

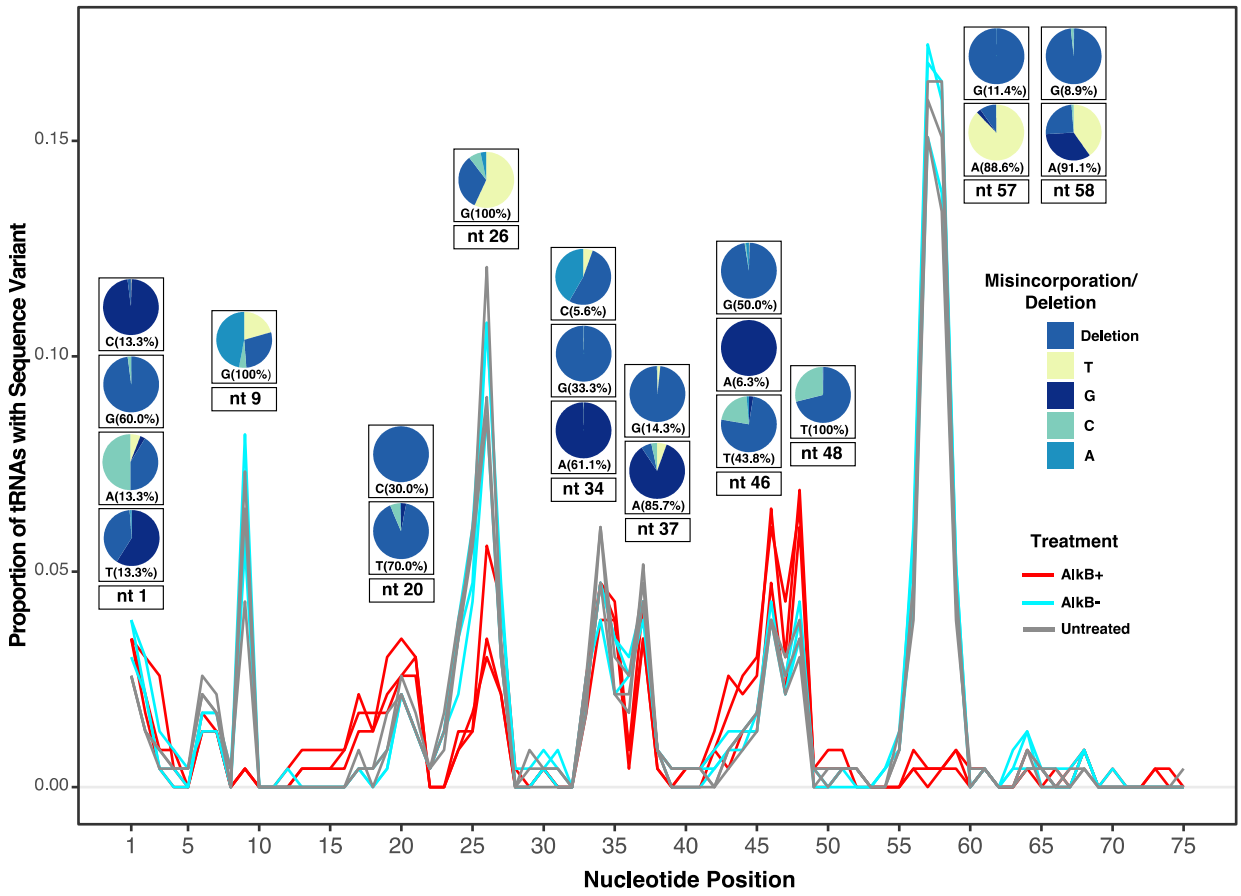

Fig. 7| A tRNA modification index showing the proportion of all *A. thaliana* tRNA reference sequences with a misincorporation/deletion at each nucleotide position. A position was considered modified if ≥30% of the mapped reads differed from the reference sequence and the sequence was detected by more than five reads. Pie charts show the identity of the misincorporated base at some of the most frequently modified positions. A separate pie chart is provided for each observed reference base at that position, and the percentage indicates what proportion of modified tRNA sequences have that reference nucleotide. All three replicates of each treatment are shown and indicated by line color. Because the y-axis is expressed as a proportion of tRNA reference sequences (and not total reads), genes such as nuclear tRNA-Pro and plastid tRNA-GlyGCC are not preferentially weighted in this analysis even though they represent the majority of reads.

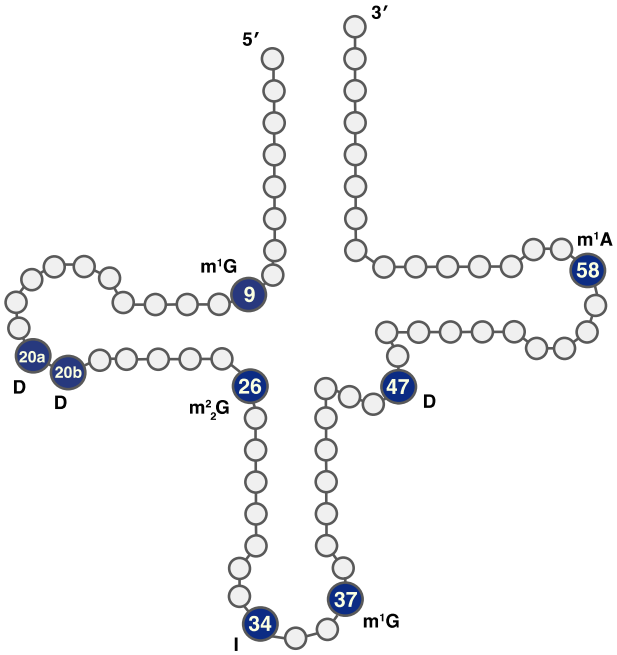

Fig. 8| tRNA positions labeled with possible modifications. Positions with strong misincorporation signals are indicated with a known modification at that position in other species. D: dihydrouridine; I: inosine; m^1A: N1-methyladenosine; m^1G: N1-methylguanosine; $m^2{}_2G$: N2,N2-dimethylguanosine.

REFERENCES

1. Motorin Y, Muller S, Behm-Ansmant I, et al. Identification of modified residues in RNAs by reverse transcription-based methods. Methods Enzymol. 2007;425:21-53.
2. Wilusz JE. Removing roadblocks to deep sequencing of modified RNAs. Nat Methods. 2015 Sep;12(9):821-2.
3. Shigematsu M, Honda S, Loher P, et al. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. Nucleic Acids Research. 2017 May 19;45(9).
4. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57-63.
5. Trewick SC, Henshaw TF, Hausinger RP, et al. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature. 2002 Sep 12;419(6903):174-8.
6. Zheng G, Qin Y, Clark WC, et al. Efficient and quantitative high-throughput tRNA sequencing. Nat Methods. 2015 Sep;12(9):835-837.
7. Cozen AE, Quartley E, Holmes AD, et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nature Methods. 2015 Sep;12(9):879-+.
8. Dai Q, Zheng G, Schwartz MH, et al. Selective Enzymatic Demethylation of N(2) ,N(2) -Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. Angew Chem Int Ed Engl. 2017 Apr 24;56(18):5017-5020.
9. Smith AM, Abu-Shumays R, Akeson M, et al. Capture, Unfolding, and Detection of Individual tRNA Molecules Using a Nanopore Device. Front Bioeng Biotechnol. 2015;3:91.
10. Pang YL, Abo R, Levine SS, et al. Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. Nucleic Acids Res. 2014 Dec 16;42(22):e170.
11. Zhong J, Xiao C, Gu W, et al. Transfer RNAs Mediate the Rapid Adaptation of Escherichia coli to Oxidative Stress. Plos Genet. 2015 Jun;11(6):e1005302.
12. Torres AG, Reina O, Stephan-Otto Attolini C, et al. Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. Proc Natl Acad Sci U S A. 2019 Apr 23;116(17):8451-8456.
13. Gogakos T, Brown M, Garzia A, et al. Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. Cell Rep. 2017 Aug 8;20(6):1463-1475.
14. Hoffmann A, Fallmann J, Vilardo E, et al. Accurate mapping of tRNA reads. Bioinformatics. 2018 Jul 1;34(13):2339.
15. Rojas-Benitez D, Thiaville PC, de Crecy-Lagard V, et al. The Levels of a Universally Conserved tRNA Modification Regulate Cell Growth. J Biol Chem. 2015 Jul 24;290(30):18699-707.
16. Lyons SM, Fay MM, Ivanov P. The role of RNA modifications in the regulation of tRNA cleavage. FEBS Lett. 2018 Sep;592(17):2828-2844.
17. Vermeulen A, McCallum SA, Pardi A. Comparison of the global structure and dynamics of native and unmodified tRNAval. Biochemistry. 2005 Apr 26;44(16):6024-33.

18. Phizicky EM, Alfonzo JD. Do all modifications benefit all tRNAs? FEBS Lett. 2010 Jan 21;584(2):265-71.
19. Sampson JR, Uhlenbeck OC. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. Proc Natl Acad Sci U S A. 1988 Feb;85(4):1033-7.
20. Hou YM, Gamper H, Yang W. Post-transcriptional modifications to tRNA--a response to the genetic code degeneracy. RNA. 2015 Apr;21(4):642-4.
21. Urbonavicius J, Qian Q, Durand JM, et al. Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J. 2001 Sep 3;20(17):4863-73.
22. Huber SM, Leonardi A, Dedon PC, et al. The Versatile Roles of the tRNA Epitranscriptome during Cellular Responses to Toxic Exposures and Environmental Stress. Toxics. 2019 Mar 25;7(1).
23. Phizicky EM, Hopper AK. tRNA processing, modification, and subcellular dynamics: past, present, and future. RNA. 2015 Apr;21(4):483-5.
24. Kuchino Y, Hanyu N, Nishimura S. Analysis of modified nucleosides and nucleotide sequence of tRNA. Methods Enzymol. 1987;155:379-96.
25. Ross R, Cao X, Yu N, et al. Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. Methods. 2016 Sep 1;107:73-8.
26. Kowalak JA, Pomerantz SC, Crain PF, et al. A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. Nucleic Acids Res. 1993 Sep 25;21(19):4577-85.
27. Youvan DC, Hearst JE. Reverse transcriptase pauses at N2-methylguanine during in vitro transcription of Escherichia coli 16S ribosomal RNA. Proc Natl Acad Sci U S A. 1979 Aug;76(8):3751-4.
28. Schwartz S, Motorin Y. Next-generation sequencing technologies for detection of modified nucleotides in RNAs. RNA Biol. 2017 Sep 2;14(9):1124-1137.
29. Clark WC, Evans ME, Dominissini D, et al. tRNA base methylation identification and quantification via high-throughput sequencing. RNA. 2016 Nov;22(11):1771-1784.
30. Kietrys AM, Velema WA, Kool ET. Fingerprints of Modified RNA Bases from Deep Sequencing Profiles. J Am Chem Soc. 2017 Nov 29;139(47):17074-17081.
31. Motorin Y, Helm M. Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies. Genes-Basel. 2019 Jan;10(1).
32. Kimura S, Dedon PC, Waldor MK. Surveying the landscape of tRNA modifications by combining tRNA sequencing and RNA mass spectrometry. bioRxiv. 2019:723049.
33. Hummel G, Warren J, Drouard L. The multi-faceted regulation of nuclear tRNA gene transcription. IUBMB Life. 2019 Aug;71(8):1099-1108.
34. Bermudez-Santana C, Attolini CS, Kirsten T, et al. Genomic organization of eukaryotic tRNAs. BMC Genomics. 2010 Apr 28;11:270.
35. Goodarzi H, Nguyen HCB, Zhang S, et al. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell. 2016 Jun 2;165(6):1416-1427.
36. Kondo K, Makovec B, Waterston RH, et al. Genetic and molecular analysis of eight tRNA(Trp) amber suppressors in Caenorhabditis elegans. J Mol Biol. 1990 Sep 5;215(1):7-19.

37. Theologis A, Ecker JR, Palm CJ, et al. Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. Nature. 2000 Dec 14;408(6814):816-20.
38. Cognat V, Pawlak G, Duchene AM, et al. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res. 2013 Jan;41(Database issue):D273-9.
39. Soprano AS, Smetana JHC, Benedetti CE. Regulation of tRNA biogenesis in plants and its link to plant growth and response to pathogens. Biochim Biophys Acta Gene Regul Mech. 2018 Apr;1861(4):344-353.
40. Park EJ, Kim TH. Fine-Tuning of Gene Expression by tRNA-Derived Fragments during Abiotic Stress Signal Transduction. Int J Mol Sci. 2018 Feb 8;19(2).
41. Cognat V, Morelle G, Megel C, et al. The nuclear and organellar tRNA-derived RNA fragment population in Arabidopsis thaliana is highly dynamic. Nucleic Acids Res. 2017 Apr 7;45(6):3460-3472.
42. Salinas-Giegé T, Giegé R, Giegé P. tRNA Biology in Mitochondria. Int J Mol Sci. 2015 Mar;16(3):4518-4559.
43. Machnicka MA, Olchowik A, Grosjean H, et al. Distribution and frequencies of post-transcriptional modifications in tRNAs. RNA Biol. 2014;11(12):1619-29.
44. Iida K, Jin H, Zhu JK. Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in Arabidopsis thaliana. BMC Genomics. 2009 Apr 9;10:155.
45. Chen HC, Viry-Moussaid M, Dietrich A, et al. Evolution of a mitochondrial tRNA PHE gene in A. thaliana: import of cytosolic tRNA PHE into mitochondria. Biochem Biophys Res Commun. 1997 Aug 18;237(2):432-7.
46. Duchêne AM, Drouard L. The chloroplast-derived trnW and trnM-e genes are not expressed in Arabidopsis mitochondria. Biochem Biophys Res Commun. 2001 Aug 3;285(5):1213-6.
47. Novoa EM, Pavon-Eternod M, Pan T, et al. A role for tRNA modifications in genome structure and codon usage. Cell. 2012 Mar 30;149(1):202-13.
48. Fuchs RT, Sun Z, Zhuang F, et al. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. PLoS One. 2015;10(5):e0126049.
49. Vandivier LE, Anderson ZD, Gregory BD. HAMR: High-Throughput Annotation of Modified Ribonucleotides. Methods Mol Biol. 2019;1870:51-67.
50. Potapov V, Fu X, Dai N, et al. Base modifications affecting RNA polymerase and reverse transcriptase fidelity. Nucleic Acids Res. 2018 Jun 20;46(11):5753-5763.
51. Banerjee R, Chen S, Dare K, et al. tRNAs: cellular barcodes for amino acids. FEBS Lett. 2010 Jan 21;584(2):387-95.
52. Hopper AK, Phizicky EM. tRNA transfers to the limelight. Genes Dev. 2003 Jan 15;17(2):162-80.
53. Phizicky EM, Hopper AK. tRNA biology charges to the front. Genes Dev. 2010 Sep 1;24(17):1832-60.
54. Kirchner S, Ignatova Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nat Rev Genet. 2015 Feb;16(2):98-112.
55. Wilusz JE. Controlling translation via modulation of tRNA levels. Wiley Interdiscip Rev RNA. 2015 Jul-Aug;6(4):453-70.

56. Huang SQ, Sun B, Xiong ZP, et al. The dysregulation of tRNAs and tRNA derivatives in cancer. J Exp Clin Cancer Res. 2018 May 9;37(1):101.
57. Chatrchyan S, Khachatryan V, Sirunyan AM, et al. Determination of jet energy calibration and transverse momentum resolution in CMS. J Instrum. 2011 Nov;6.
58. Raina M, Ibba M. tRNAs as regulators of biological processes. Front Genet. 2014;5:171.
59. Selden RF, Steinmetz A, McIntosh L, et al. Transfer RNA genes ofZea mays chloroplast DNA. Plant Mol Biol. 1983 May;2(3):141-53.
60. Bergmann P, Seyer P, Burkard G, et al. Mapping of transfer RNA genes on tobacco chloroplast DNA. Plant Mol Biol. 1984 Jan;3(1):29-36.
61. Mubumbila M, Crouse EJ, Weil JH. Transfer RNAs and tRNA genes of Vicia faba chloroplasts. Curr Genet. 1984 Jul;8(5):379-85.
62. Jacob D, Thuring K, Galliot A, et al. Absolute Quantification of Noncoding RNA by Microscale Thermophoresis. Angew Chem Int Ed Engl. 2019 Jul 8;58(28):9565-9569.
63. Hauenschild R, Tserovski L, Schmid K, et al. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. Nucleic Acids Res. 2015 Nov 16;43(20):9950-64.
64. Telonis AG, Loher P, Magee R, et al. tRNA Fragments Show Intertwining with mRNAs of Specific Repeat Content and Have Links to Disparities. Cancer Res. 2019 Jun 15;79(12):3034-3049.
65. Stupar RM, Lilly JW, Town CD, et al. Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci U S A. 2001 Apr 24;98(9):5099-103.
66. Huang CY, Ayliffe MA, Timmis JN. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature. 2003 Mar 6;422(6927):72-6.
67. Telonis AG, Kirino Y, Rigoutsos I. Mitochondrial tRNA-lookalikes in nuclear chromosomes: could they be functional? RNA Biol. 2015;12(4):375-80.
68. Telonis AG, Loher P, Kirino Y, et al. Nuclear and mitochondrial tRNA-lookalikes in the human genome. Front Genet. 2014;5:344.
69. Sequeira-Mendes J, Araguez I, Peiro R, et al. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell. 2014 Jun;26(6):2351-2366.
70. Jain M, Olsen HE, Paten B, et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016 Nov 25;17(1):239.
71. Pan T. Modifications and functional genomics of human transfer RNA. Cell Res. 2018 Apr;28(4):395-404.
72. Tserovski L, Marchand V, Hauenschild R, et al. High-throughput sequencing for 1-methyladenosine (m(1)A) mapping in RNA. Methods. 2016 Sep 1;107:110-21.
73. Ma X, Si F, Liu X, et al. PRMdb: a repository of predicted RNA modifications in plants. Plant and Cell Physiology. 2020;In press.
74. Jackman JE, Alfonzo JD. Transfer RNA modifications: nature's combinatorial chemistry playground. Wiley Interdiscip Rev RNA. 2013 Jan-Feb;4(1):35-48.
75. Hendrickson TL. Recognizing the D-loop of transfer RNAs. Proc Natl Acad Sci U S A. 2001 Nov 20;98(24):13473-5.

76. Torres AG, Pineyro D, Rodriguez-Escriba M, et al. Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. Nucleic Acids Res. 2015 May 26;43(10):5145-57.
77. Arimbasseri AG, Blewett NH, Iben JR, et al. RNA Polymerase III Output Is Functionally Linked to tRNA Dimethyl-G26 Modification. Plos Genet. 2015 Dec;11(12):e1005671.
78. Xing F, Hiley SL, Hughes TR, et al. The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. J Biol Chem. 2004 Apr 23;279(17):17850-60.
79. Lorenz C, Lunse CE, Morl M. tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomolecules. 2017 Apr 4;7(2).
80. Geslain R, Pan T. Functional analysis of human tRNA isodecoders. J Mol Biol. 2010 Feb 26;396(3):821-31.
81. Kashdan MA, Dudock BS. The gene for a spinach chloroplast isoleucine tRNA has a methionine anticodon. J Biol Chem. 1982 Oct 10;257(19):11191-4.
82. Alkatib S, Fleischmann TT, Scharff LB, et al. Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. Nucleic Acids Res. 2012 Aug;40(14):6713-24.
83. Su D, Chan CT, Gu C, et al. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. Nat Protoc. 2014 Apr;9(4):828-41.
84. Boccaletto P, Machnicka MA, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. 2018 Jan 4;46(D1):D303-D307.
85. Pereira M, Francisco S, Varanda AS, et al. Impact of tRNA Modifications and tRNA-Modifying Enzymes on Proteostasis and Human Disease. Int J Mol Sci. 2018 Nov 24;19(12).
86. de Crecy-Lagard V, Boccaletto P, Mangleburg CG, et al. Matching tRNA modifications in humans to their known and predicted enzymes. Nucleic Acids Res. 2019 Mar 18;47(5):2143-2159.
87. Bjork GR, Jacobsson K, Nilsson K, et al. A primordial tRNA modification required for the evolution of life? EMBO J. 2001 Jan 15;20(1-2):231-9.
88. Ebhardt HA, Tsang HH, Dai DC, et al. Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. Nucleic Acids Res. 2009 May;37(8):2461-70.
89. Werner S, Schmidt L, Marchand V, et al. Machine learning of reverse transcription signatures of variegated polymerases allows mapping and discrimination of methylated purines in limited transcriptomes. Nucleic Acids Res. 2020 Apr 17;48(7):3734-3746.
90. Henley RY, Carson S, Wanunu M. Studies of RNA Sequence and Structure Using Nanopores. Prog Mol Biol Transl Sci. 2016;139:73-99.
91. Poodari VC. Direct RNA Sequencing of E. coli initiator tRNA Using the MinION Sequencing Platform: UC Santa Cruz; 2019.
92. Onanuga K, Begley TJ, Chen AA, et al. tRNA Modification Detection Using Graphene Nanopores: A Simulation Study. Biomolecules. 2017 Sep;7(3).
93. Jordon-Thaden IE, Chanderbali AS, Gitzendanner MA, et al. Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. Appl Plant Sci. 2015 May;3(5).

94. Chen F, Tang Q, Bian K, et al. Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. Chem Res Toxicol. 2016 Apr 18;29(4):687-93.
95. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-12.
96. Bushnell B, Rood J, Singer E. BBMerge - Accurate paired shotgun read merging via overlap. Plos One. 2017 Oct 26;12(10).
97. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017 Sep 6.
98. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40.