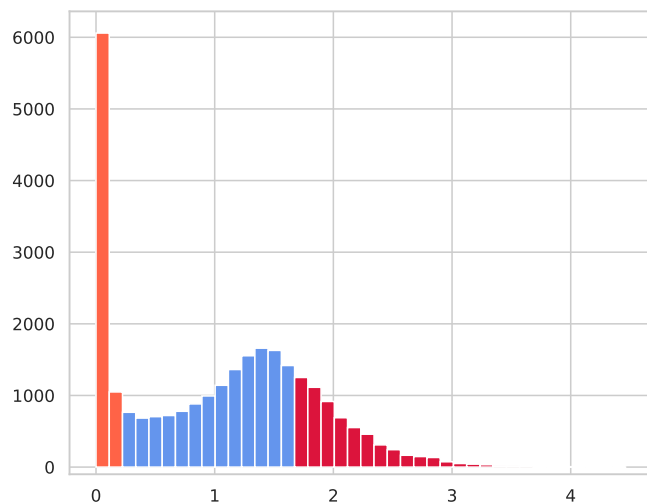
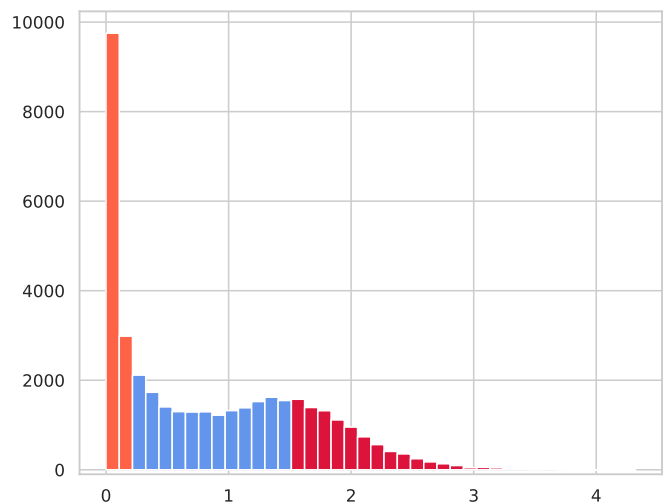
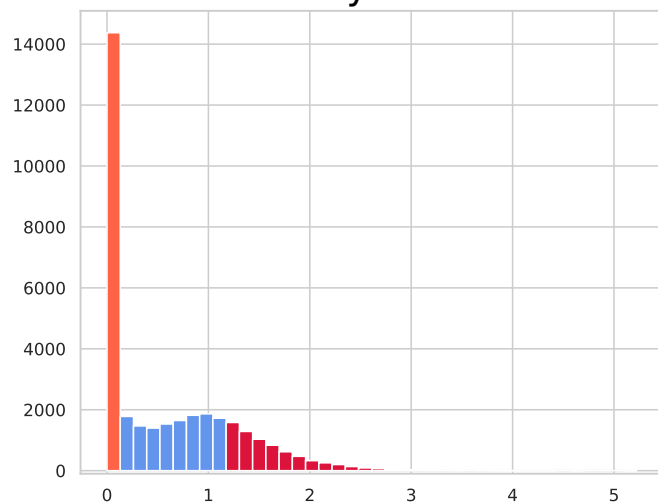
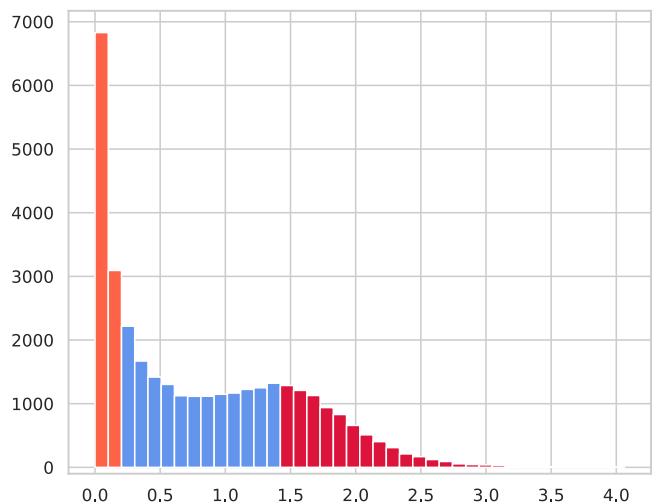


A. tha*Z. may**S. lyc**S. bic*

Expression level

low

medium

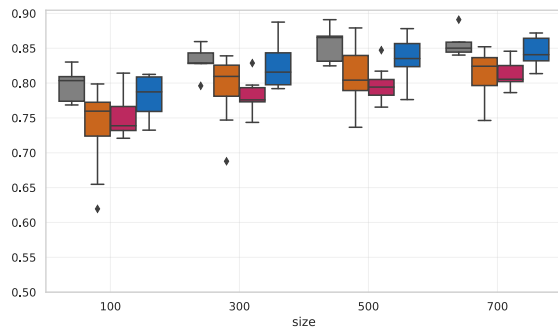
high

Supplementary Figure 1 - Short read transcriptome samples prepared from leafs were mapped to for the four reference species *A. thaliana*, *S. lycopersicum*, *S. bicolor*, and *Z. mays*. respectively. This produced normalized expression counts in transcripts per million (TPM). For each gene, we applied a logarithmic transformation to the maximum TPM value across samples to represent the expression level. Genes were categorized into low (dark orange), medium (blue), high (red) expression classes based on the lower and upper 25% quantiles of the $\log_{10}(\text{maxTPM}+1)$ distributions. The threshold values for leaf transcript profiles of *A. thaliana*, *S. lycopersicum*, *S. bicolor*, and *Z. mays* were 0.199, 0.000, 0.153, and 0.113 $\log_{10}(\text{maxTPM}+1)$ for the lower and 1.621, 1.051, 1.389, and 1.465 $\log_{10}(\text{maxTPM}+1)$ for the higher quartile, respectively (Supplementary Data 1**).**

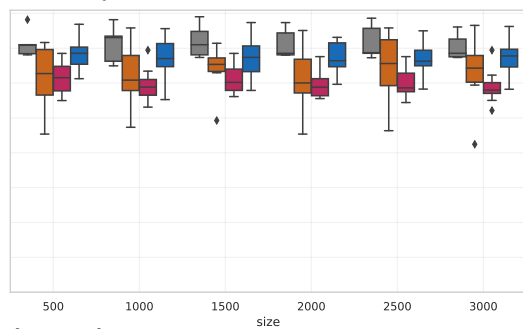
a

accuracy

UTR



promoter and terminator

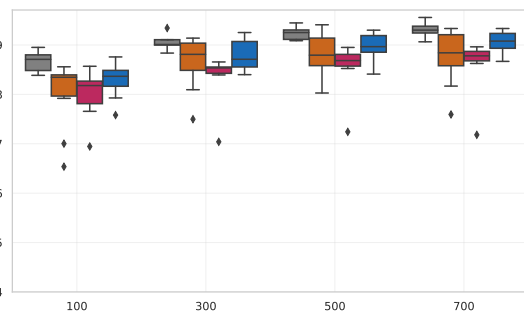


sequence length

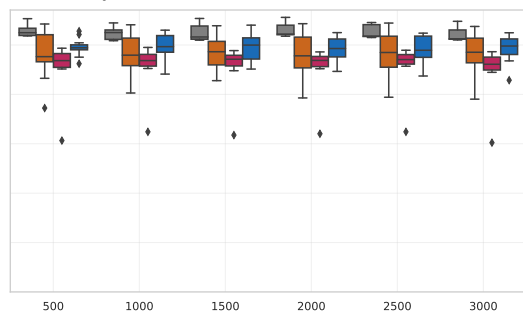
b

auROC

UTR



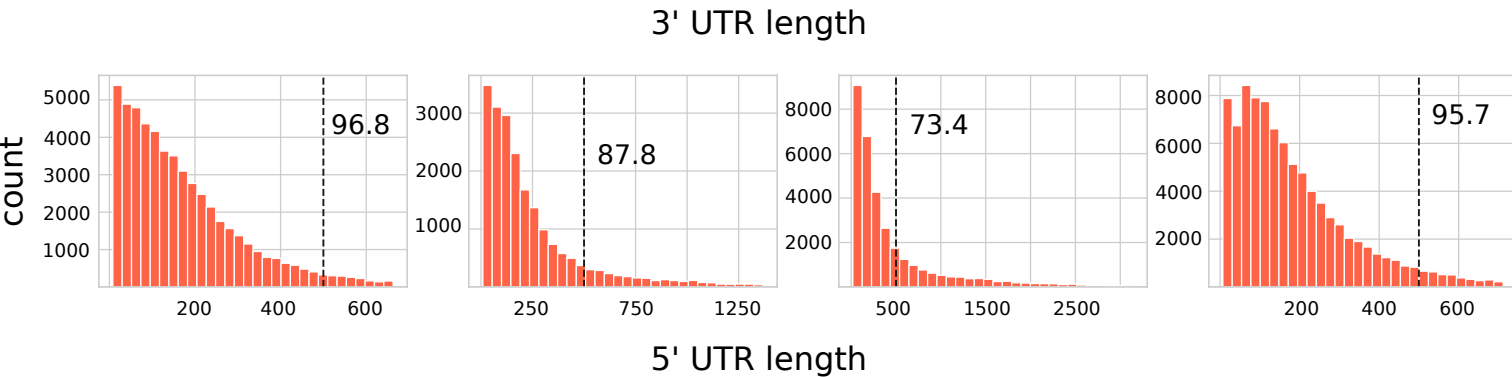
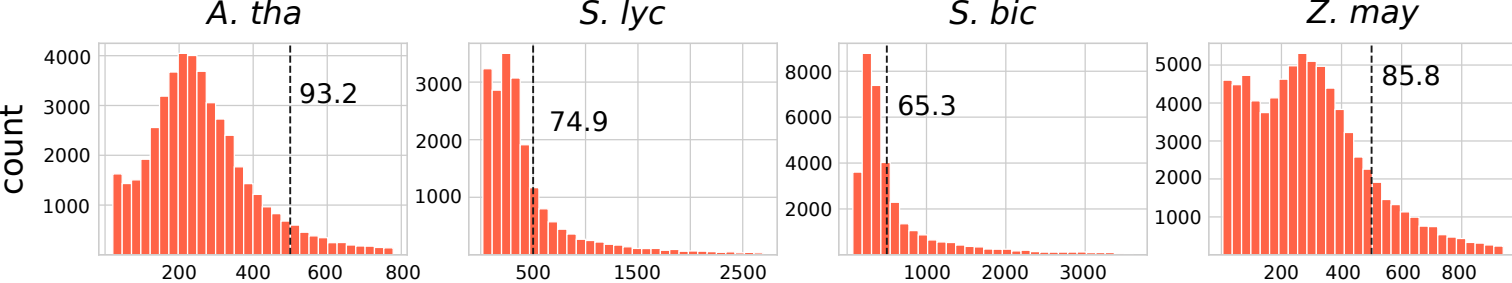
promoter and terminator



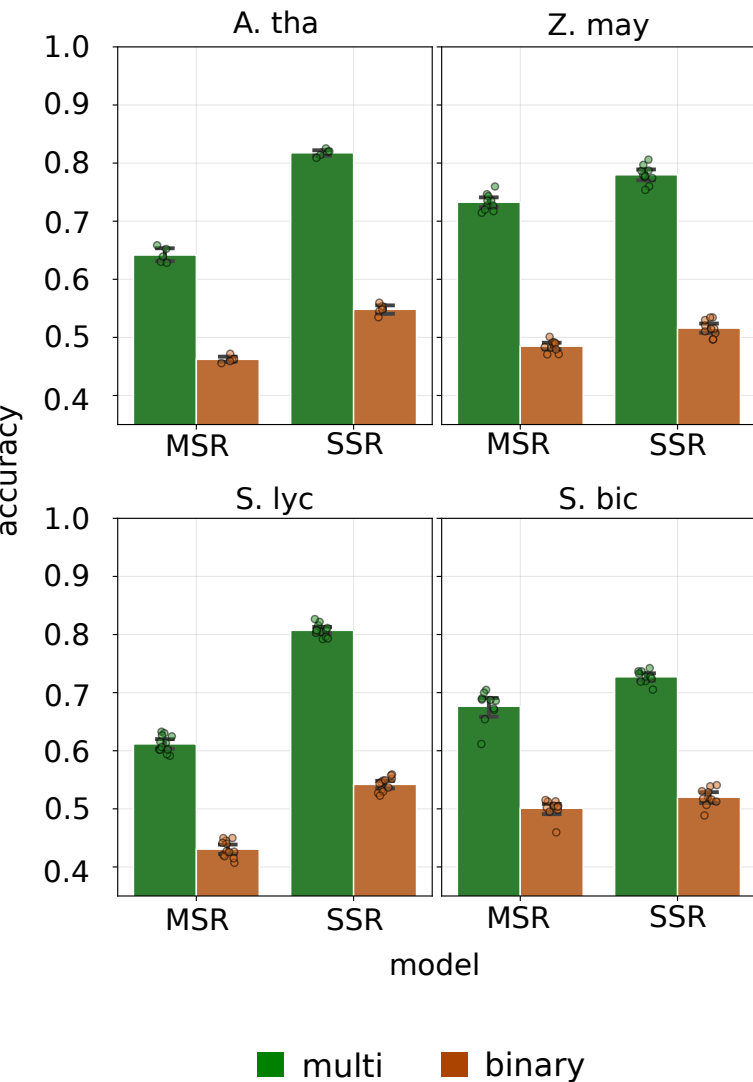
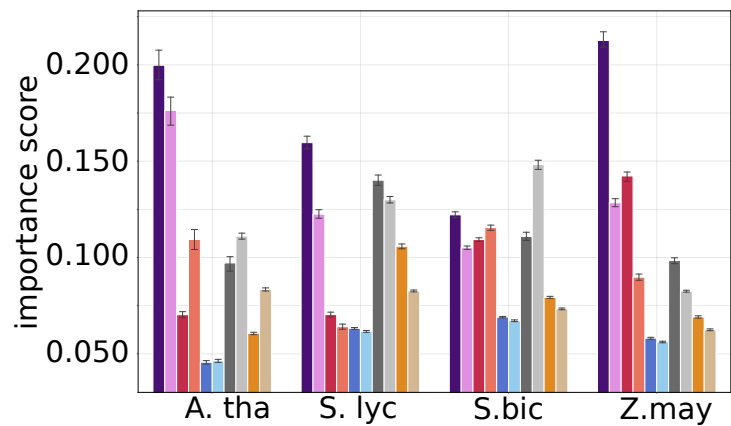
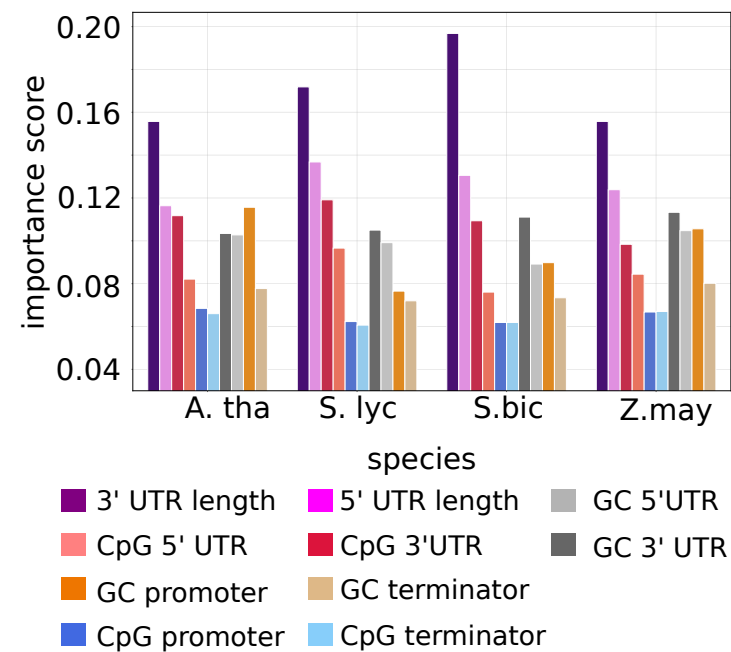
sequence length

*A. tha**S. bic**S. lyc**Z. may*

Supplementary Figure 2 - The comparison of model performance for different promoter, terminator and UTR lengths. Lengths of the 5' and 3' UTRs were varied from 100 to 700 nt while keeping the terminator and promoter lengths constant at 1000 nt. Alternatively, the 5' and 3' UTR lengths were kept constant at 500 nt while varying the promoter and terminator lengths from 500 to 3000 nt. Model performance in accuracy (a) and area under the receiver operating characteristic curve (b) was recorded for varied UTR (left) and varied promoter and terminator lengths (right). Boxplot depicting sample characteristics using the 25th, 50th (median), and 75th percentiles along with the interquartile range, representing the central 50% of the data. Whiskers extend from the minimum to maximum values, showcasing the spread of the dataset.



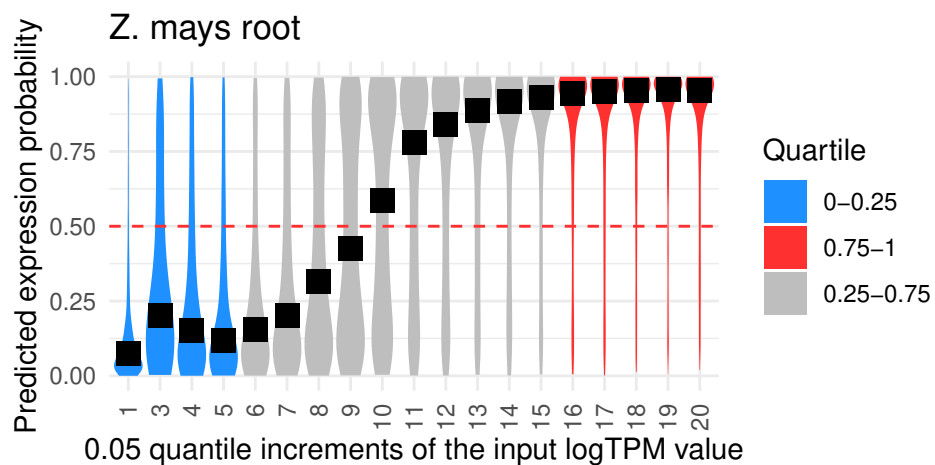
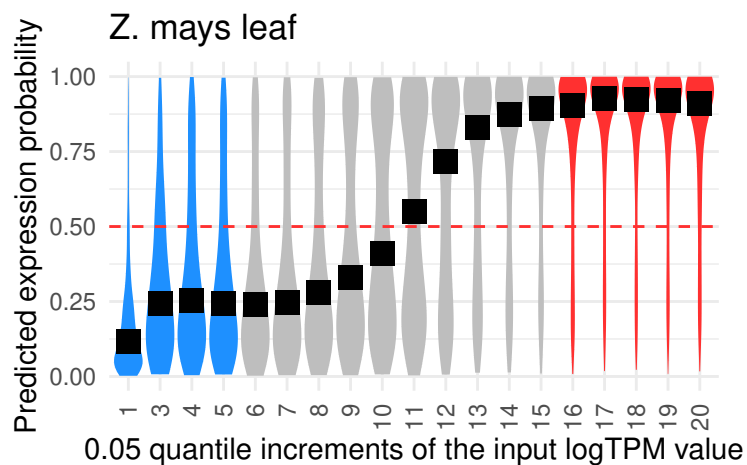
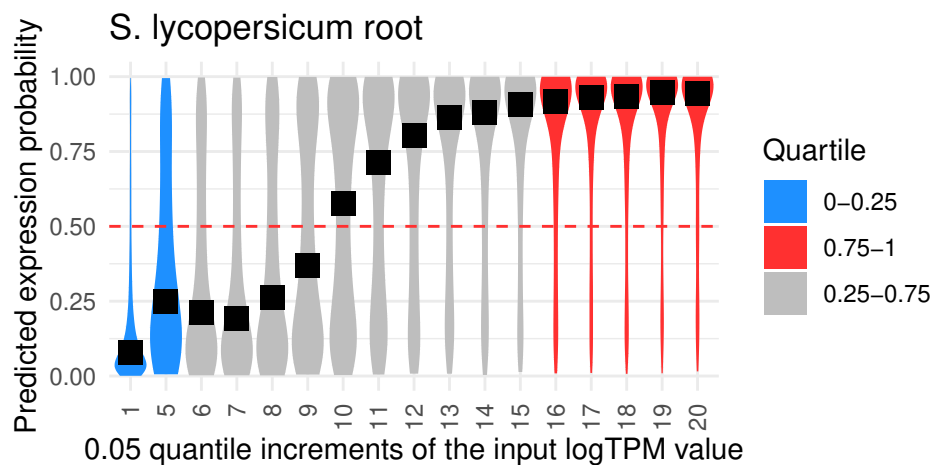
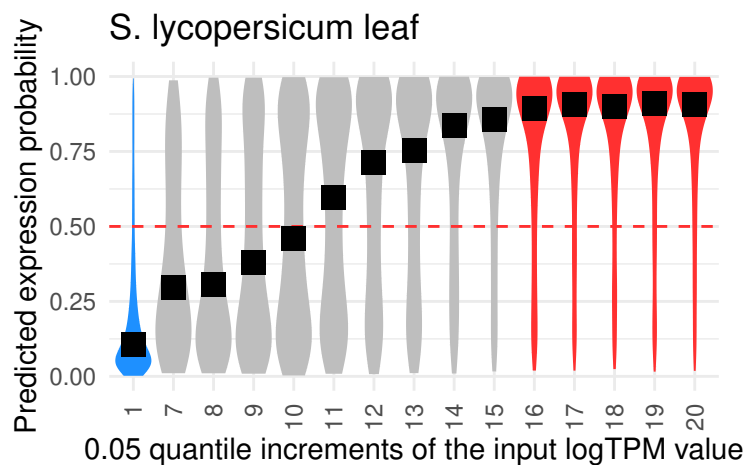
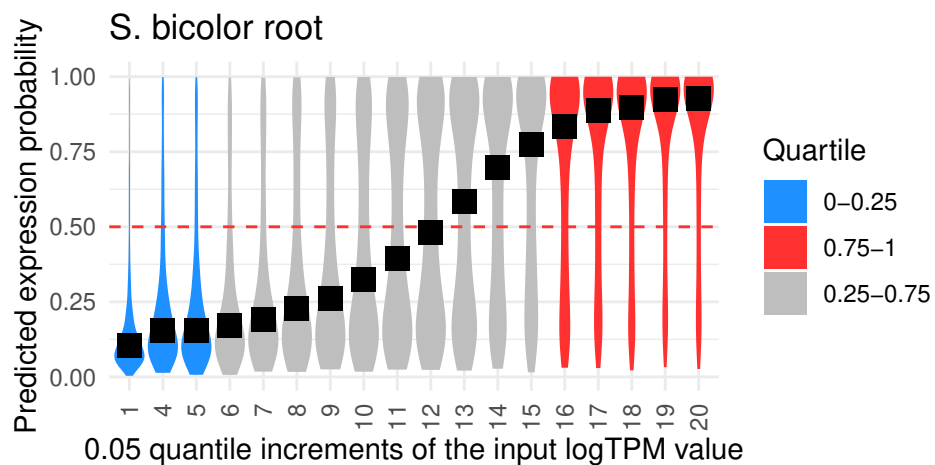
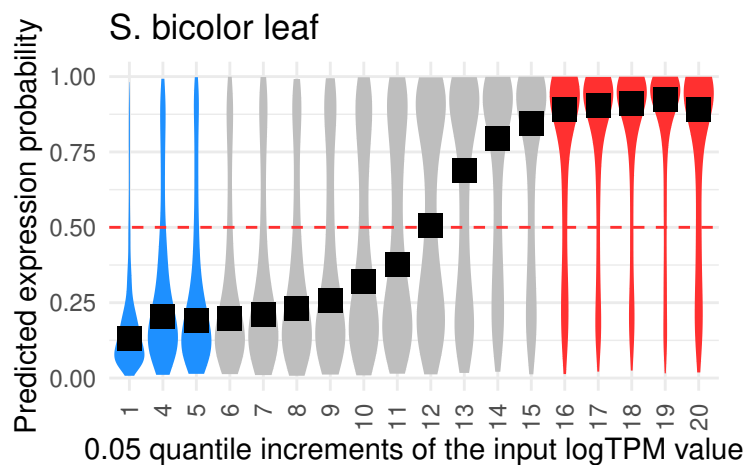
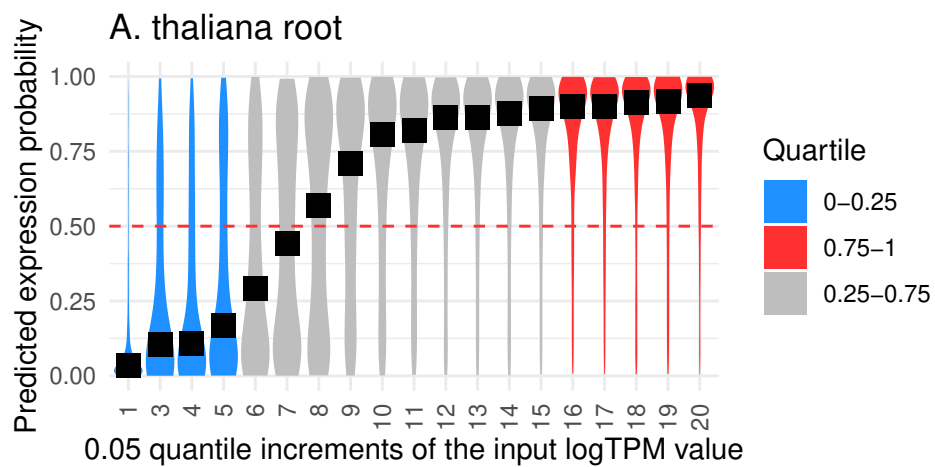
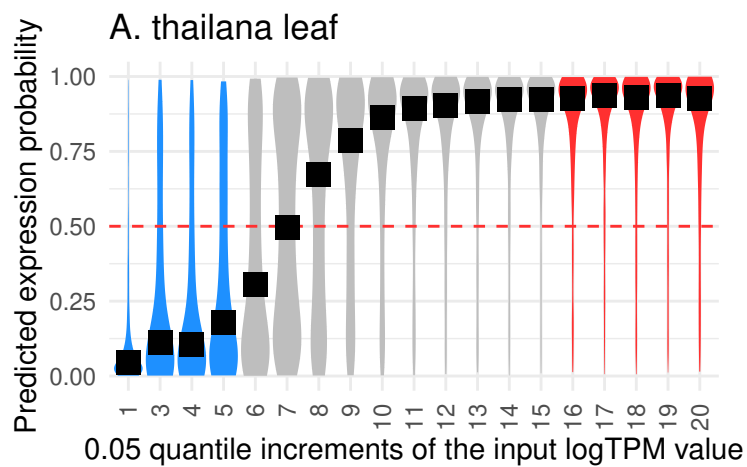
Supplementary Figure 3 - The distribution of the UTR lengths of three prime and five prime ends for the four reference species *A. thaliana* (A. tha.), *S. lycopersicum* (S. lyc.), *S. bicolor* (S. bic.) and *Z. mays* (Z. may). Histograms show the counts of genes with distinct UTR lengths (red bars). There were 65.3 to 96.8 % of *S. bicolor* three prime UTR to *A. thaliana* five prime UTR shorter than 500 bp, indicated by the dashed line.

a**b****c**

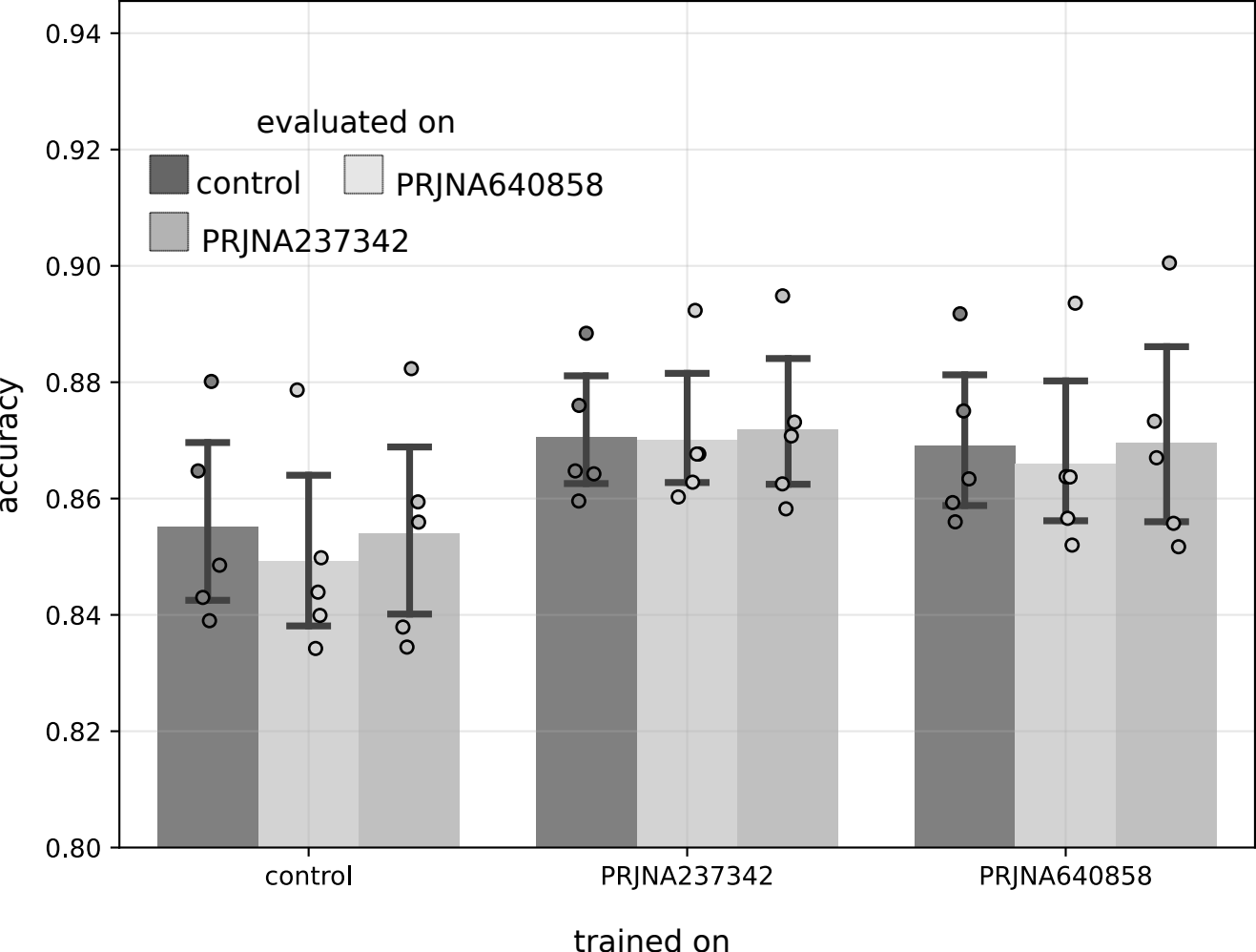
Supplementary Figure 4 - Random forest models for the four reference plants trained on generic features generated based on the reference gene models. The models achieved high accuracies on test performance for SSR models but not the MSR setup, but outperformed prediction of multiple transcript level classes.

(a) Accuracy of random forest classifiers for binary (low-high) and multi (low-medium-high) classification models, are shown for the reference species *A. thaliana*, *Z. mays*, *S. lycopersicum*, and *S. bicolor*. Individual data points are depicted as circles ($n = 5, 12, 10$ and 10 for *A. thaliana*, *S. lycopersicum*, *S. bicolor*, and *Z. mays*). The error bars show the 95% confidence intervals and levels.

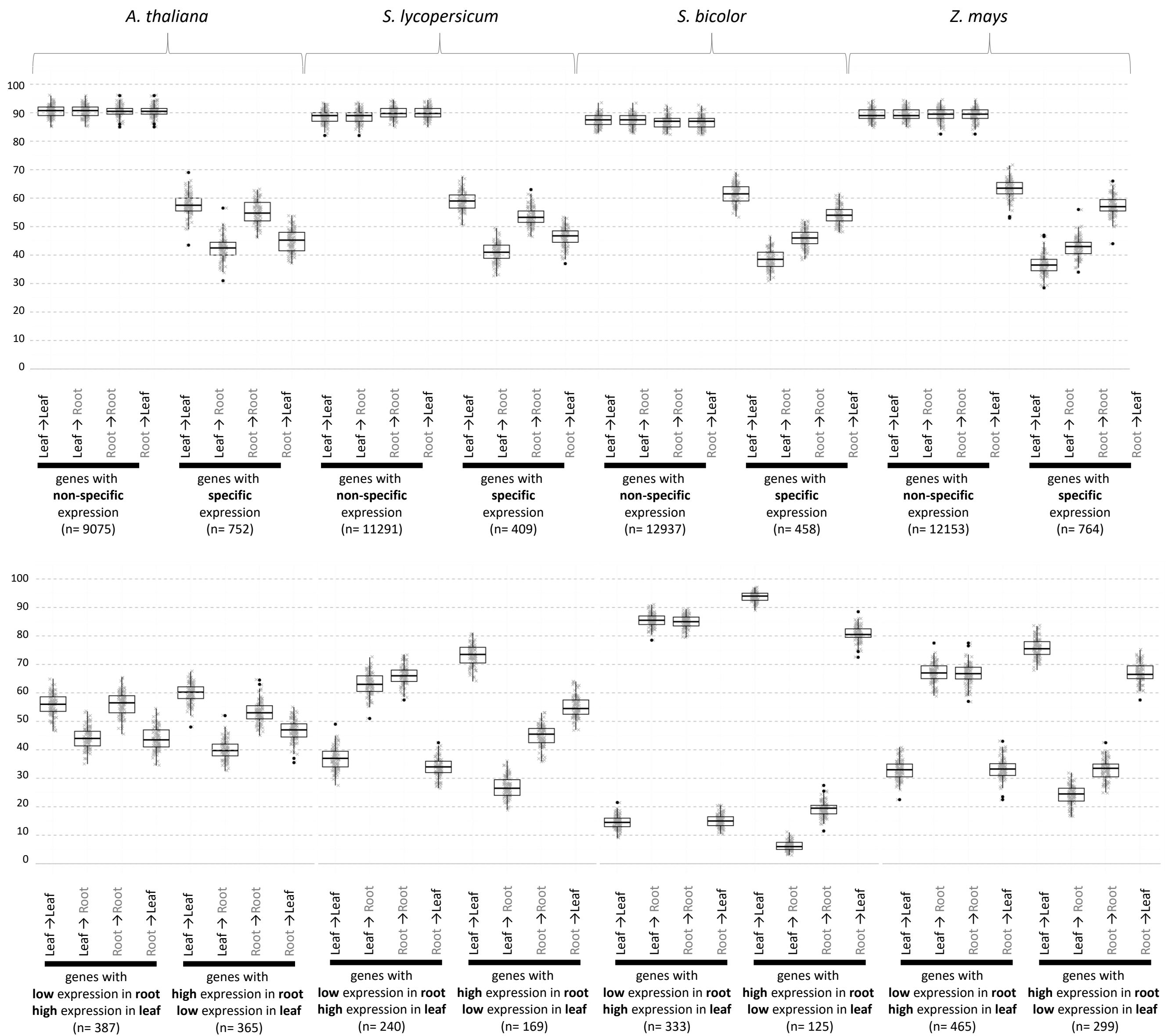
(b) Impurity-based feature importance from SSR random forest classifiers. random forest classifiers trained (c) Impurity-based feature importance from MSR random forest classifiers. The error bars show the 95% confidence intervals and levels.



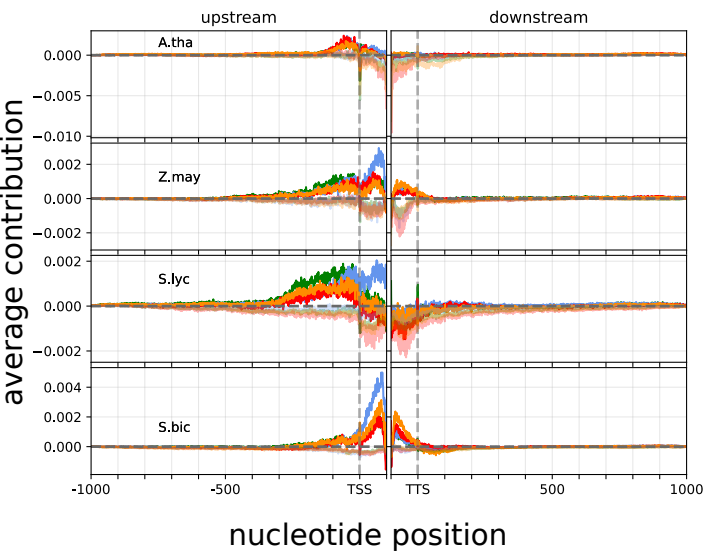
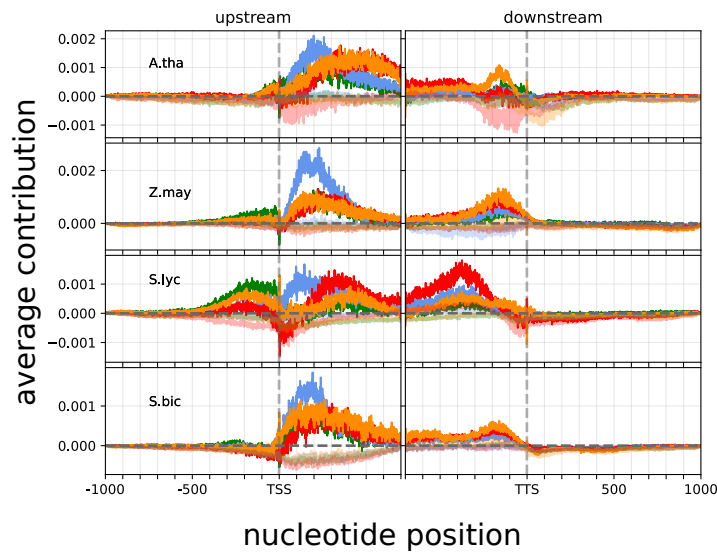
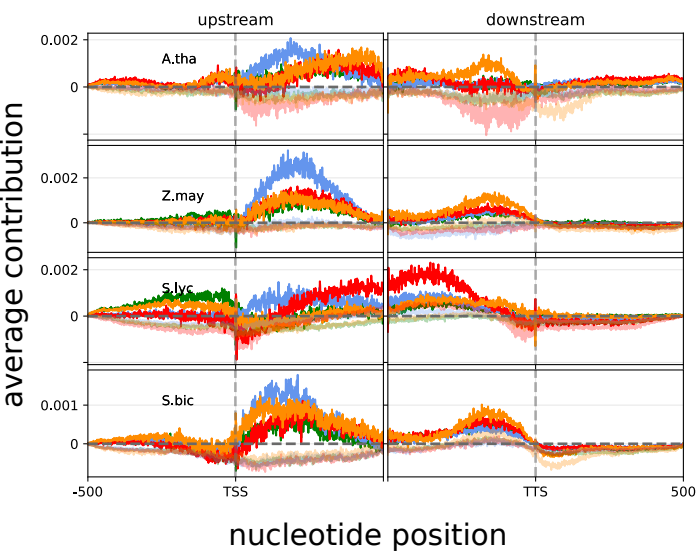
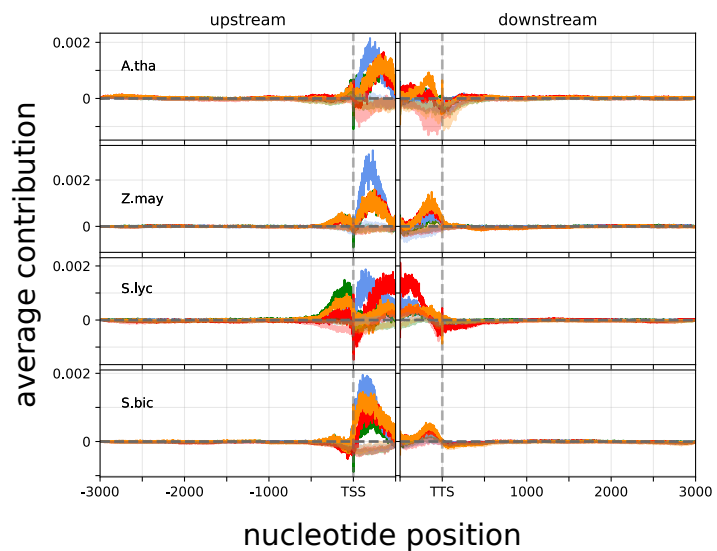
Supplementary Figure 5 - Distribution of prediction probabilities for low and high levels of gene expression of SSR models. Violin plots display the predicted probabilities with mean (black square) for each consecutive 5% quantile of the $\log_{10}\text{maxTPM}+1$ values for each of the four species *A. thaliana*, *S. bicolor*, *S. lycopersicum* and *Z. mays* leaf and root samples, respectively.



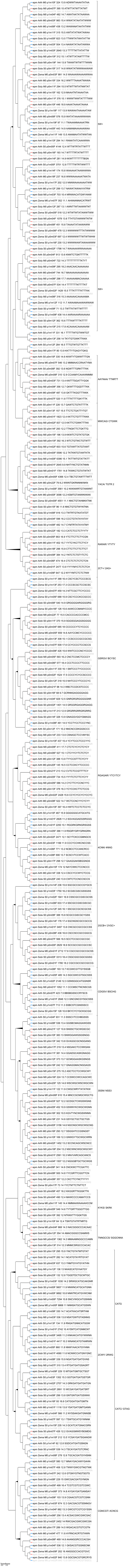
Supplementary Figure 6 - Cross prediction performance for SSR models trained by different experimental data. SSR models for *A. thaliana* were generated from RNAseq experiment profiles PRJEB32665, PRJNA237342 and PRJNA640858 and individual predictive performance was cross evaluated to estimate the training strategy sensitivity to the effects of technical or experimental variation. Individual data points (n=5) are depicted as circles. The error bars show the 95% confidence intervals.



Supplementary Figure 7 - Cross-validation of deep learning SSR models of leaf and root tissue was conducted for four plant species: *A. thaliana*, *S. lycopersicum*, *S. bicolor*, and *Z. mays*. To assess the predictive accuracy of these models for gene expression levels, we partitioned the test sets into genes with tissue-specific and non-specific expression patterns. Tissue specificity was determined based on gene expression classification; genes exhibiting opposite expression patterns in root and leaf tissues were considered tissue specific. We performed 1,000 bootstrap iterations to calculate average accuracies, accounting for varying sample sizes. Additionally, we further subdivided genes with tissue-specific expression to investigate tissue and expression class-specific effects on model performance, despite very small sample sizes. The boxplots depict summary statistics of individual average accuracies (gray crosses) per test condition randomly sampled by bootstrapping using the 25th, 50th (median), and 75th percentiles along with the interquartile range, representing the central 50% of the data. Whiskers extend from the minimum to maximum values, showcasing the spread of the dataset and outliers are depicted as ovals.

a**b****c****d**

Supplementary Figure 8 - Saliency maps for model training on different input sequence lengths. The model are interpreted by the calculation of DeepLift importance score for models trained with shorter UTRs of 100 nt (a), longer UTRs of 700 nt (b), shorter promoter-terminators of 500 nt (c) and longer promoter-terminators of 3000 nt (d).



Supplementary Figure 9 - Dendrogram of 520 EPMs of *A. thaliana*, *S. bicolor*, *S. lycopersicum* and *Z. mays* from SSR and MSR leaf models. EPMs were clustered by the Smith-Waterman algorithm. The resulting clades of EPMs with similar consensus sequences were named after alignment (**Supplementary Data 5**). EPMs associated with low (blue leaf-tips) and high rates (pink leaf tips) of gene expression find similar motifs across different models from SSR leaf (S0) and MSR leaf (M0). EPMs in reverse orientation are displayed gray.

Supplementary Figure 10 – The comparison of EPMs from *A. thaliana* and *Z. mays* SSR leaf and root models. EPMs of *A. thaliana* (a) and *Z. mays* (b) SSR leaf (green) and root (brown) models exhibit a high degree of similarity. With over 80% of motifs having significant similarity towards one or more EPMs comparing both leaf and root EPMs with over 95% (p-value > 0.05). This similarity is determined through Pearson Correlation Coefficient.

(c) Smith-waterman-clustered dendrogram for *A. thaliana* SSR leaf and root model EPMs with similar clades found in the comparison of SSR and MSR leaf model EPMs (**Supplementary Figure 9**).

Supplementary Note 1 - Identification and characterization of predictive gene model features. While the interpretation of the models revealed a conserved pattern of associated saliency scores and the presence of specific EPMs, the contribution of general features of gene models is difficult to infer from the CNN model. Therefore, we generated ten generic features per gene and trained random forest (RF) classifiers for the SSR and MSR training strategy that were chosen arbitrarily by contribution scoring. For the 5' and 3' regions five generic features each were chosen: UTR length, CpG UTR content, CpG promoter region content, GC UTR content and GC promoter region content. Remarkably, our RF-SSR classifiers achieved accuracy values of 81.76%, 80.72%, 72.72%, and 77.98% in *A. thaliana*, *S. lycopersicum*, *S. bicolor*, and *Zea mays*, respectively (**Supplementary Figure 4**). In contrast, a lower performance for the RF-MSR models with accuracy values of 64.16%, 61.17%, 67.65% and 73.25% was observed for the species, respectively (**Supplementary Figure 4**). The impurity-based feature contributions calculated for the RF-SSR model indicated that the UTR-based features were the most important for classification with the lengths of the 3' UTRs being the most important feature across all species (**Supplementary Figure 4**). 5'UTR length showed higher importance for genes of *A. thaliana*, while 3'UTR GC content was highly predictive in genes of *S. lycopersicum*. The impurity-based contributions for the RF-MSR model also showed the importance of the UTR-based features, with the length of the 3'UTR being the most important feature (**Supplementary Figure 4**).