May 2021

A Tissue Specific Transcriptomic, Proteomic and Phospho-proteomic Atlas of the Translational Machinery of Arabidopsis thaliana

Abdullah Salim
asalim1@vols.utk.edu

81. A Tissue Specific Transcriptomic, Proteomic and Phospho-proteomic Atlas of the Translational Machinery of Arabidopsis thaliana

Abdullah Salim[1], Ricardo Urquidi Camacho[2], Albrecht von Arnim[1,2]

[1]Department of Biochemistry, Cellular and Molecular Biology, [2]Graduate School of Genome Science and Technology

Abstract

Gene expression encompasses the flow of genetic information from DNA to mRNA (transcription) and from mRNA to protein (translation) along with the regulatory mechanisms underlying these processes. Omics technologies offer a powerful toolset with which to study gene expression at each of these stages. A recently published dataset integrating transcriptomic, proteomic and phospho-proteomic measurements from 30 Arabidopsis thaliana tissues provides a unique resource to explore gene expression. The translational machinery (the ribosome, and its initiation, elongation, and termination factors) are a core component in gene expression. Defects in translation can be lethal or lead to major developmental defects and hold keys to better understand crop immunity and yield. In this work we have developed a suite of visualizations for genes found in the translational machinery for each omics dataset. Furthermore, we examine ribosomal heterogeneity across tissues and estimate the stoichiometry of subcomponents of the translational machinery. We also describe phosphorylation patterns across tissues, elucidating potential translational regulatory patterns across tissues. The conclusions of this project can be utilized to enhance understanding of the regulation of translational machinery and generate new hypotheses that can be tested in a wet lab setting.

Measuring Gene Expression

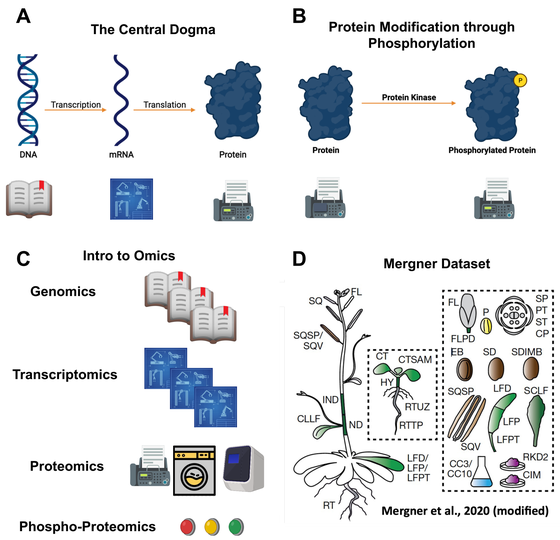


Figure 1. An overview and introduction to gene expression, omics, and the Mergner Dataset. **A)** The Central Dogma, which refers to the flow of genetic information from DNA to mRNA and from mRNA to protein, is a key part of gene expression and protein synthesis. **B)** Once a protein is synthesized, it can undergo post-translational modifications (PTMs), such as acylation, glycosylation, and phosphorylation (pictured). PTMs can alter the functional state of a protein. **C)** Omics offer us an opportunity to study bulk gene expression at different stages of the Central Dogma. **D)** The Mergner Dataset measures the transcriptome, proteome, and phospho-proteome for 30 tissues in *Arabidopsis thaliana*.

Research Goals

- Create a visual atlas of the translational machinery
- Study the heterogeneity of translational machinery components
- Study differences in phospho-regulation of the translational machinery across tissues

A Visual Atlas: An Overview

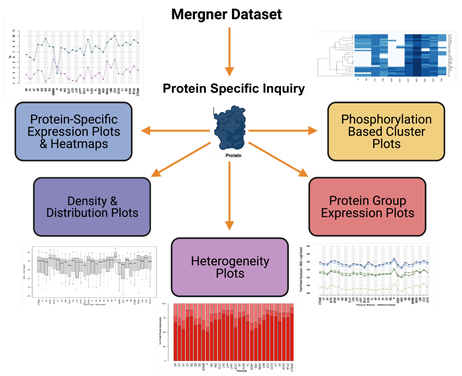


Figure 2. An overview of the different types of visualizations that were developed to study gene expression in different tissues, such as pollen (pictured). Included are protein-specific expression plots and heatmaps to visualize mRNA and protein expression. Density and distribution plots analyze the quality of expression (not shown on poster). Heterogeneity plots show the contribution of protein paralog expression to total expression of a functional protein group. Protein group expression plots explore the balance of translational machinery components. Phosphorylation cluster plots visualize and analyze phosphorylation events.

Protein Specific Visualizations

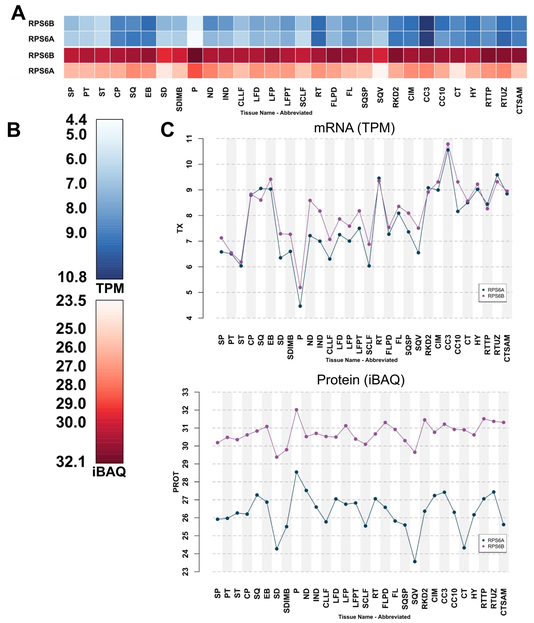
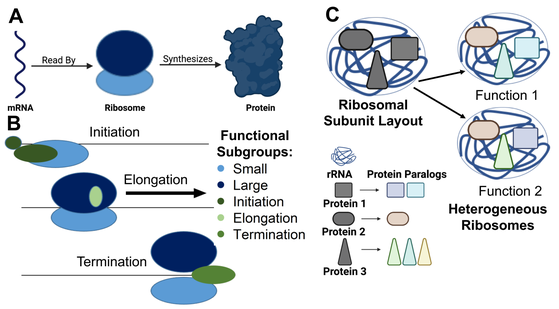


Figure 3. **A)** A heatmap visualizing mRNA expression (blue) and protein expression (red) for two proteins, RPS6A and RPS6B, across the 30 tissues analyzed. **B)** The scale used to interpret the heatmaps – intensity of expression is based on the intensity of the hue. mRNA is expressed in transcripts per million (TPM). Protein expression is expressed in terms of intensity based absolute quantification (iBAQ). **C)** Line plots serve as an alternative way of visualizing expression across the 30 tissues, allowing for a more quantitative visualization of expression.

❖ **The creation of a visual atlas facilitates the study of mRNA and protein expression of different genes across different tissues.**

Tissue Key: SP – Sepal, PT – Petal, ST – Stamen, CP – Carpel, SQ – Silique, EB – Embryo, SD – Seed, SDIMB – Seed Imbibed, P – Pollen, ND – Node, IND – Internode, CLLF – Cauline Leaf, LFD – Leaf Distal, LFP – Leaf Proximal, LFPT – Leaf Petiole – SCLF – Senescent Leaf, RT – Root, FLPD – Flower Pedicel, FL – Flower, SQSP – Silique Septum, SQV - Silique Valve, RKD2 – Callus, CIM – Initial Cell Culture, CC10 – Late Cell Culture, CT – Cotyledon, HY – Hypocotyl, RTTP – Root Tip, RTTUZ – Root Tip Upper Zone, CTSAM – Cotyledon and Shoot Apical Meristem

Ribosomal Heterogeneity



D Paralogs in each functional subgroup (mRNA/Protein)

	Small	Large	Initiation	Elongation	Termination
Total	99/99	134/134	73/73	16/16	6/6
All Tissues	74/51	112/74	55/47	14/11	5/5
At Least 1 Tissue	25/12	19/11	17/8	2/2	1/1
None	0/36	3/49	1/18	0/3	0/0

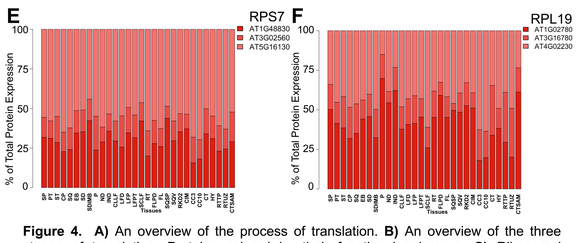


Figure 4. **A)** An overview of the process of translation. **B)** An overview of the three stages of translation. Proteins colored by their functional subgroup. **C)** Ribosomal subunits are composed of rRNA and ribosomal proteins (RP). Ribosomal heterogeneity may arise from the usage of different RP paralogs. **D)** Number of paralogs present in each tissue for each functional subgroup. **E)** and **F)** Paralog composition percentage for E) RPS7 protein expression and F) RPL19 protein expression.

❖ **Most paralogs are expressed as mRNA. Several paralogs are missing protein expression.**
❖ **Some RP show heterogeneity in the use of their protein paralogs.**

Translational Coordination

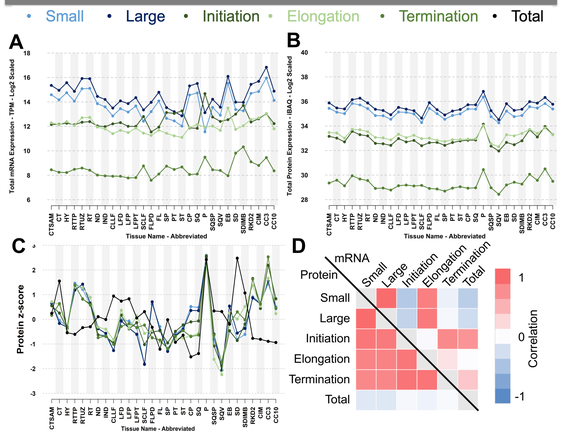


Figure 5. **A)** mRNA expression for each functional subgroup. **B)** Protein expression for each functional subgroup. **C)** Standardized (z-score) protein expression for each functional subgroup compared to total protein expression (TPE). **D)** Correlation between functional subgroups and TPE for mRNA and protein expression.

❖ **Gene expression within functional subgroups remains relatively constant throughout all 30 tissues.**
❖ **Functional subgroups are highly correlated at the protein level but less so at the mRNA level.**

Phosphorylation Patterns of the Translational Machinery

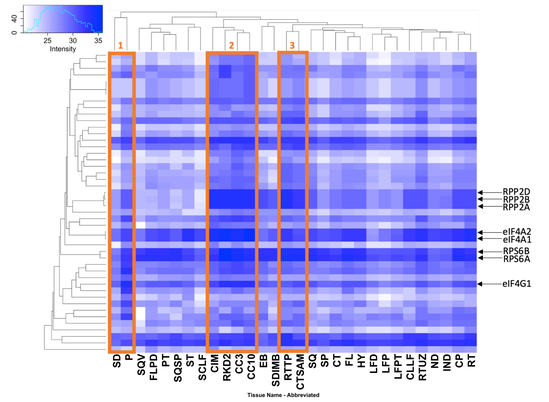


Figure 6. Euclidean clustering of the sum of phospho-peptide intensities of the translational machinery. Three tissue clusters were identified as having strong patterns of phosphorylation tied to biological function: 1.) Protein and Seed, 2.) Calli and Cell Culture, and 3.) Meristems. All of these tissues have high protein production requirements. Proteins highlighted are well known to be phosphorylated.

❖ **Differential phosphorylation patterns are observed across tissues. Potentially indicating tissue specific roles.**

Figure 7. Heatmaps showing tissue-specific phosphorylation intensity for RPS6A (left) and RPS6B (right) for each phosphorylation site (p-site).

❖ **S237, S240 are well known conserved p-sites of unknown function throughout eukarya. Their differential phosphorylation across tissues may help elucidate their function.**
❖ **S229, S231 p-sites that are phosphorylated in tissues with a higher proportion stem cells or with high protein production requirements.**

Conclusions

❖ The main value of this work is in describing patterns and generating hypotheses for wet lab experimentation through the use of our visual atlas.
❖ Some ribosomal proteins may show heterogeneity in their paralog expression. However, more careful analysis will be needed for paralog disambiguation at the proteome level.
❖ Functional subgroups of the translational machinery are highly correlated at the protein level but not at the mRNA level, indicating non-transcriptional gene regulation.
❖ Higher phosphorylation patterns are found in tissues reliant on higher protein production.

Future Work:

❖ Paralog level analysis of phosphorylation and development of better methods for clustering phosphorylation data.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Acknowledgements

Mergner Dataset Source:
Mergner, J., Frejno, M., List, M. *et al.* Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* **579,** 409–414 (2020). https://doi.org/10.1038/s41586-020-2094-2

Images produced with: BioRender.com

Special thanks to Dr. Albrecht von Arnim, Ricardo Urquidi Camacho, and the Von Arnim Lab for their mentorship and support throughout this project.