Chapter 23

Multi-omics methods applied to flower development

Raquel Álvarez-Urdiola[a], José Tomás Matus[a,b] and José Luis Riechmann[a,c,*]

[a] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Edifici CRAG, Campus UAB, 08193 Cerdanyola del Vallès, Barcelona, Spain.

[b] Institute for Integrative Systems Biology (I2SysBio), Universitat de València-CSIC, Paterna, 46908, Valencia, Spain.

[c] Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

* Corresponding author: joseluis.riechmann@cragenomica.es

i. Running Head

Flower development multi-omics.

ii. Summary/Abstract

Developmental processes in multicellular organisms depend on the proficiency of cells to orchestrate different gene expression programs. Over the past years, several studies of reproductive organ development have considered genomic analyses of transcription factors and global gene expression changes, modelling complex gene regulatory networks. Nevertheless, the dynamic view of developmental processes requires, as well, the study of the proteome in its expression, complexity, and relationship with the transcriptome. In this chapter we describe a dual extraction method -for protein and RNA- for the characterization of genome expression at proteome level and its correlation to transcript expression data. We also present a shotgun proteomic procedure (LC-MS/MS) followed by a pipeline for the imputation of missing values in mass spectrometry results.

iii. Key Words

Protein extraction, RNA extraction, proteomics, transcriptomics, flower development, LC-MS/MS, Arabidopsis.

1. Introduction

The capacity of cells to orchestrate different gene expression programs is crucial for developmental processes in multicellular organisms, and it is hardwired and encoded in the genome in the form of *cis*-regulatory sequences that interact with transcription factors, co-regulators, and other types of regulatory proteins or RNAs, as well as of epigenetic marks, altogether determining when, where, and how genes are expressed. For the past twenty years, the exponential advances in technologies and informatics tools for generating and processing large biological datasets (omics) have added new approaches to development studies in plants. Through the use of genomics and transcriptomics (in particular RNA-Seq, ChIP-Seq, and other high-throughput sequencing-derived methods), the hierarchical levels of plant genetic and molecular organization are being described in detail. In particular, several studies of reproductive organ development have considered genome-wide analyses of transcription factor DNA-binding and global gene expression changes (e.g., ***(1–5)***) and modelled complex gene regulatory networks (reviewed in ***(6–9)***). Even so, a global and comprehensive view of developmental processes would also benefit from the characterization of the corresponding proteome.

The analysis of the proteome of eukaryotic cells is challenging due to the substantial diversity in the properties of the individual proteins that compose it (e.g., abundance, stability, molecular weight, structure, hydrophobicity, hydrophilicity, post-translational modifications -PTMs-, and so on). Nevertheless, along with an enhancement of throughput, sensitivity, and resolution of analytical technologies in MS, computational methods have been developed focusing on the identification and quantification of proteins in complex samples ***(10–13)***. In plants, MS-based proteomics approaches have been applied for the measurement of differential protein expression or the detection of

PTMs *(*e.g., ***14,15)*** in different tissues and biological processes (reviewed in ***(13)***). Deep proteome studies have led to the development of proteome atlases of the major plant organs for different plant species ***(16–21)***. Besides, cell type-specific proteome studies are crucial for a better understanding of the unique biological functions and properties of individual cell types in a tissue ***(22)***, as well as subcellular plant proteomics and predictions ***(23–25)***. As the proteome is in constant flux, several proteome studies are based on temporal series during developmental processes, or stress responses ***(26-29)***.

Furthermore, results from more than one type of omics can be matched in order to obtain deeper insight into biological processes ***(16,30–33)***. These integration studies are usually referred as multi-omics, trans-omics, or integrated omics in current literature. Quantitative proteomics allows to study at a genome-wide level the correlation between mRNA expression levels and the abundance of the corresponding proteins (reviewed in ***(34,35)***), an issue that has been extensively studied in different species and processes during the past few years. For instance, in plants combined transcriptome-proteome analyses have already been used to study petal shape ***(36)***, carotenoid synthesis ***(37)***, photoperiodic control of the proteome ***(38)***, or leaf development ***(39)***,as well as reproductive development; in particular, embryogenesis ***(40)***, male reproductive development ***(41–43)***, flower development in general ***(44,45)*** or focused on the functions of specific proteins ***(46)***.

In these combined studies, the interpretation of the existence, or lack thereof, of correlation between the changes in transcript dynamics and protein abundance, and its biological meaning, is still a lingering issue: numerous studies conclude that there is not a strong correlation between the levels of these macromolecules ***(41,43,47–51)***, whereas in others such correlation is more apparent ***(38–40,45)***. The lack of correlation could be in part derived from the difficulties to obtain truly comparable datasets at the transcript

and protein levels, and because the sensitivity of extraction and quantification techniques for mRNAs and proteins highly differ. However, the observed differences might also be caused by posttranslational regulation of protein levels ***(47)***, or by their different expression and degradation kinetics, as longer protein half-lives buffer changes in mRNA levels ***(48–51)***. Time-lapse studies could be an approach for addressing this gap, as successive analyses at different time points could allow the discovery of correlative behaviors of protein and mRNA levels through time ***(52,53)***.

In addition, a major concern in label-free quantitative proteomics that hinders the subsequent data analysis and its comparison with other omics data is the high rate of missing values. Three types of missing values can be defined, depending on the nature of the missingness: i) Missing Completely At Random (MCAR) and ii) Missing At Random (MAR) values, which are due to minor errors or stochastic fluctuations and to conditional dependencies, respectively; and iii) Missing Not At Random (MNAR) values, which have a targeted effect ***(54)***. Depending on the nature of these "not assigned values" (NAs), different methods can be used to impute them. As there are many types of NAs that coexist in most quantitative datasets, hybrid strategies of imputation could be a better approach ***(54,55)***.

In this chapter, we describe a protocol for common extraction of total proteins and RNA from the same Arabidopsis inflorescence samples to maximize comparability between the proteomic and transcriptomic data. We also present a shotgun proteomic procedure by liquid chromatography-tandem mass spectrometry (LC-MS/MS), and a pipeline for the imputation of missing values in the mass spectrometry results to distinguish the nature of the missingness and to treat NAs accordingly.

2. Materials

1. Mortar and pestle.

2. Liquid nitrogen.

3. Microcentrifuge tubes.

2.1 Protein extraction

1. Protein low binding tubes (2 mL).

2. Isopropanol.

3. 0.3 M Guanidine in 95% ethanol.

4. 90% ethanol.

5. SDS-PAGE 5x buffer.

6. E buffer: 125 mM Tris-HCl pH 8.8, 1 % (w/v) SDS, 10% (v/v) glycerol, 50 mM $Na_2S_2O_5$ ***(56)***.

2.2 RNA extraction

1. RNase free tubes (1.5 mL).

2. Trizol.

3. Chloroform.

4. Phenol:Chloroform:Isoamyl alcohol (25:24:1).

5. LiCl 3M.

6. 85% and 100% (v/v) ethanol.

7. DEPC water.

2.3 LC-MS/MS

1. DL-Dithiothreitol (DTT) (*see* **Note 1**).

2. Iodoacetamide.

3. Urea.

4. Ammonium bicarbonate.

5. Endoproteinase LysC.

6. Trypsin.

7. Formic acid.

8. MicroSpin C18 columns (The Nest Group, Inc).

9. Nano Trap C18 columns with an inner diameter of 100 μm packed with C18 particles of 5 μm particle size (Thermo Fisher Scientific) (Optional, depending on the setup of each laboratory).

10. Reverse-phase chromatography columns (C18, 2 μm, 15-50 cm length) (*see* **Note 2**).

11. Buffer A: 0.1% formic acid in water.

12. Buffer B: 0.1% formic acid in acetonitrile.

13. Bovine serum albumin (New England Biolabs cat # P8108S).

14. Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) (*see* **Note 3**).

15. EASY-nLC 1000 (Thermo Fisher Scientific).

3. Methods

3.1 Protein extraction

1. With a different mortar and pestle for each sample, grind the tissue (i.e., inflorescences) with liquid nitrogen until obtaining a whitish fine powder (*see* **Notes 4** and **5**).

2. Place the powder in a microcentrifuge tube (~250 mg per sample).

3. Add 1 mL of Trizol, vortex for at least 15 seconds until it is completely homogenized, and incubate on ice for 5 min. This step must be done in an extraction hood.

4. Add 200 µL of chloroform, vortex for 15 seconds, incubate on ice for 5 min, and centrifuge at 4ºC for 15 min at maximum speed (*see* **Note 6**) (**Figure 1**).

5.a. Transfer 500 – 600 µL of the top, aqueous phase into a clean microcentrifuge tube (RNase free) and add the same volume of phenol:chloroform:isoamyl alcohol, vortex for 10 seconds, incubate on ice for 5 min, and centrifuge at 4ºC for 15 min at maximum speed (to continue with RNA extraction from the sample, *see* **3.2 RNA extraction**).

5.b. Add 300 µL of ethanol 100% to the organic phase in the original microcentrifuge tube to continue with protein extraction. Incubate on ice.

6. Centrifuge for 10 min at 2000 g. Place the supernatant in a clean 2 mL microcentrifuge tube (protein low bind).

7. Add 1 mL of isopropanol and incubate at room temperature for 10 min (*see* **Note 7**).

8. Centrifuge at 4ºC for 10 min at 12000 g. Discard supernatant, which contains phenol, into a container adequate for its controlled elimination.

9. Wash by resuspending the pellet in 2 mL of a solution of 0.3 M guanidine in 95% ethanol (*see* **Note 8**).

10. Sonicate in a sonication bath for 5 min and centrifuge at 4ºC for 5 min at 8000 g.

11. Repeat the washing procedure (steps 9 and 10) twice. The obtained pellet can be stored at -20ºC for months.

12. Wash again by the same procedure (steps 9-11) with 90% ethanol.

13. Let the pellet dry for a few minutes and resuspend in an appropriate buffer (*see* **Note 9**).

14. Quantify by Bradford with 1 and 2 µL of sample. Add SDS-PAGE 5x buffer to obtain a final 1x concentration when loading the gel.

3.2 RNA extraction

1. Transfer approximately 500 µL of the top, aqueous phase after the centrifugation in Protein extraction Step 5.a to a clean microcentrifuge tube (RNase free) and add 1 volume (500 µL) of pure isopropanol. Shake and mix.

2. Incubate on ice for 15 min, centrifuge at 4ºC for 10 min at maximum speed and discard supernatant.

3. Resuspend the pellet in 750 µL of LiCl 3M, incubate on ice for 10 min and centrifuge at 4ºC for 10 min at maximum speed.

4. Discard supernatant and wash the pellet with 500 µL of ethanol 85% (v/v), vortexing gently for 10 seconds.

5. Centrifuge at 4ºC for 10 min at maximum speed and discard supernatant.

6. Let the pellet dry and resuspend in 21 µL of diethylpyrocarbonate (DEPC) treated water (*see* **Note 10**).

7. Sample quantification with NanoDrop Spectrophotometer.

3.3 LC-MS/MS

3.3.1 Sample preparation

1. Prepare or dissolve protein samples (3.1, Step 13) in 6 M Urea 200 nM ammonium bicarbonate.

2. Reduce the samples (10 µg of protein) with 30 nmol DTT at 37 ºC for 1 h.

3. Alkylate the samples (10 µg of protein) in the dark with 60 nmol of iodoacetamide at 25ºC for 30 min.

4. Dilute the protein extract to 2 M urea with 200 mM ammonium bicarbonate for digestion with endoproteinase LysC (1:10 w:w), and incubate 37ºC, overnight.

5. Dilute 2-fold with 200 mM ammonium bicarbonate for trypsin digestion (1:10 w:w), and incubate at 37ºC for 8h.

6. After digestion, add formic acid (10% of the final volume) to acidify the peptide mix.

7. Desalt the samples with MicroSpin C18 columns prior to LC-MS/MS analysis, following manufacturer's instructions.

3.3.2 Chromatographic and mass spectrometric analysis

1. Load the peptides onto the analytical column (C18, 2 µm, 15-50 cm length).

2. Separation of the peptides by reverse-phase chromatography with the corresponding columns.

3. Chromatographic gradients start at 93% buffer A and 7% buffer B with a flow rate of 250 nl/min for 5 minutes and gradually increase 65% buffer A and 35% buffer B in 60 min.

4. After each analysis, wash the column for 15 min with 10% buffer A and 90% buffer B.

5. Peptide eluates are dried in a vacuum centrifuge, and resuspended with Buffer A at a final concentration of 1 µg/µl prior to analysis by LC-MS/MS.

6. Operate the mass spectrometer to acquire peptide spectra (*see* **Note 11**).

3.3.3 Data analysis

1. Search the acquired spectra against the desired peptide database (*see* **Note 12**), plus a list of common contaminants (suggested: ***(57)***), and all the corresponding decoy entries.

2. Set the parameters accordingly to the experimental and mass spectrometric settings and, if appropriate, select variable post-translational modifications to be detected (*see* **Note 13**.

3. Determine the protein abundance estimation ***(58,59)***.

4. Add the information to the appropriate repositories (*see* **Note 14**).

3.3.4. Treatment of missing values and data imputation

1. Missing values should first be classified as M(C)AR or MNAR depending on their nature. For instance, for a given protein, if the data from all replicates of the same condition or time point show NAs, probably they are MNAR missing values, whereas if there is only one missing value out of four replicates, it is probably a MAR. Other cases may be more difficult to classify as M(C)AR or MNAR, for instance if there are two NAs out of four replicates. In those instances, other parameters can be considered, for

example the presence or absence of NAs in the adjacent time-points (in a time-course experiment) or in the most similar samples in the experiment.

2. Discard all proteins with MNARs or MARs in every sample.

3. Replace MNARs by the minimum of detection of the dataset (Deterministic Minimum Imputation method ***(60)***).

4. Estimate the remaining MARs and MCARs by other imputation method (e.g., k-Nearest Neighbor (kNN) imputation ***(61)***).

3.3.5. Example: treatment of missing values in a time series experiment

This missing value classification and data imputation approach can be readily used in, for instance, time course developmental studies ***(1,62)***, as illustrated in Figure 2 as an example. In this case, the data processing pipeline consisted on:

1. Classification of each time-point (day) for each protein depending on its number of NAs (number of replicates with missing values at a certain time-point) and the number of NAs of its immediately adjacent days (neighbors).

a. Neighbors are considered as:

- Unreliable Neighbor: Over 50% NAs.
- Reliable Neighbor: Up to 50% NAs (included).

b. Initial and final time-points are considered as:

- Reliably Undetected: 100% NAs (MNARs).
- Unreliably Undetected: Over 50% NAs (included) (unclear MNARs) + Unreliable Neighbor.

- Unreliably Detected: Over 50% NAs (included) (unclear MARs) + Reliable Neighbor.
- Reliably Detected: Up to 35% NAs (MARs).

c. Intermediate time-points are considered as:

- Reliably Undetected: 100% NAs + Unreliable Neighbors (MNARs).
- Unreliably Undetected: Over 50% NAs (included) + Unreliable Neighbors (probably MNARs).
- Unreliably Detected: Over 50% NAs (included) + Reliable Neighbors (probably MARs).
- Reliably Detected: Up to 35% NAs (MARs).

2. Replace Reliably Undetected time-points by the minimum of detection of the dataset (Deterministic Minimum Imputation method ***(60)***).

3. Replace Unreliably Undetected time-points by NAs in all replicates.

4. Discard all proteins which are Reliably or Unreliably Undetected in every time-point.

5. Estimate the remaining NAs by k-Nearest Neighbor (kNN) imputation (k = 10) ***(61)***.

4. Notes

1. Reagents for LC-MS/MS can be obtained from several suppliers. As an example, we list here the specific products we use: Urea (GE Healthcare; Sigma-Aldrich, P/N 17-1319-01); Ammonium bicarbonate (BioUltra, ≥99.5% (T); Sigma-Aldrich, P/N 09830); Iodoacetamide (BioUltra; Sigma-Aldrich, P/N I1149); DL-Dithiothreitol (for electrophoresis, ≥99%; Sigma-Aldrich, P/N D9163); Formic acid for analysis EMSURE® (ACS Reag. Merck, P/N 1.00264.0100); Sequencing Grade Modified

Trypsin (Promega, P/N V5111); Lysyl Endopeptidase (Wako Chemicals GmbH, P/N 129-02541).

2. Suitable reverse-phase chromatography columns are, for instance: 25-cm columns with an inner diameter of 75 μm, packed with 1.9 μm C18 particles (Nikkyo Technos Co.); and 50-cm columns with an inner diameter of 75 μm, packed with 2 μm C18 particles (EASY-Column, Thermo Fisher Scientific, ES903).

3. This is just a concrete example of a "modern high-resolution mass spectrometer"; other instruments could be used.

4. For sample collection, to reduce sample contamination with human proteins (i.e., keratins, collagen, etc.), make sure to always use nitrile gloves (instead of latex) and laboratory coats. Pipets, materials, and solutions exclusively used for proteomics. Take precaution to avoid hair contamination. If flower organs or tissues are going to be dissected, cool tweezers and any other sampling instrument with liquid nitrogen.

5. If samples are grown in petri dishes (e.g., Arabidopsis seedlings) discard white clots which correspond to agar.

6. Three phases are formed, the aqueous phase contains RNA (~550 μL, transparent), the interphase, DNA (white), and the organic phase, proteins and lipids (~450 μL, pink) (**Figure 1**).

7. It is possible to stop the protocol here and store the samples at -20ºC for a few days.

8. Use a pipette crushing against the bottom of the tube and leave in a colloidal suspension as thin as possible.

9. Resuspend final proteins in acetonitrile, acetic or formic acid, depending on the analysis protocol. For Western Blot, use E buffer ***(56)***. The buffer volume should be chosen depending on the desired protein concentrations, varying from 20 to 50 µL.

10. Use high pure water, reagents, and products.

11. 1 to 2 µg of peptides are loaded onto an analytical column (25-cm C18 2µm particle size) using an autosampler device (e.g., EASY nLC 1000, Thermo Fisher Scientific) and the peptides are then separated by reverse-phase chromatography using a water-acetonitril chromatographic gradient. Modern high-resolution mass spectrometers are recommended for data acquisition (e.g., Orbitrap or qTOF). The mass spectrometer is operated in data-dependent acquisition (DDA) mode, in which a full MS scan is recorded in each cycle, followed by the fragmentation of the 10-30 most intense precursor ions to obtain the fragment ion spectra.

12. The results may vary significantly depending on the characteristics of the reference database for peptide identification. It is possible to use public repositories of proteins for the different organisms or to design a specific database.

13. Once the database has been constructed, the raw LC-MS/MS data needs to be interpreted using a database search engine (such as SEQUEST ***(63)***, Mascot ***(64)***, Phenyx ***(65)***, X! Tandem ***(66)***, OMSSA ***(67)***, pFind ***(68)***, InsPecT ***(69)***, ByOnic ***(70)***, Comet ***(71)***, MS-GF+ ***(72)***, MaxQuant ***(73)*** or MSTracer ***(74)***). As example, the Mascot search engine (v2.6) can be used, using the search parameters accordingly to the experimental and mass spectrometry settings. For peptide identification a precursor ion mass tolerance below 10-20 ppm is recommended, whereas the fragment ion mass tolerance can go from 10-20 ppm for high-resolution mass analyzers (Orbitrap, TOF) to 0.5 Da if a linear ion trap is used for the analysis of the tandem mass spectra. Common

peptide modifications such as oxidation of methionine and N-terminal protein acetylation are used as variable modifications. False discovery rate (FDR) in peptide identification is set to a maximum of 1%.

14. Share data and results in a public repository. Data sharing in the public domain is the standard for omics research and a requirement for publication. For proteomics, the Proteomics IDEntifications (PRIDE) database (https://www.ebi.ac.uk/pride/) at the European Bioinformatics Institute (EMBL-EBI, Hinxton, Cambridge, UK) has enabled public data deposition of MS data since 2004 and its archival component has become the largest repository for proteomics data sharing worldwide ***(75)***. The PRIDE database provides access to most of the experimental proteomics data described in MS-related scientific publications.

Acknowledgements

Work in the authors' laboratory was supported by grant BFU2014-58289-P (funded by MICIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe") and by grant 2017SGR718 (from the Agencia de Gestió d'Ajuts Universitaris I de Recerca) to JLR, and by institutional grant SEV-2015-0533 (funded by MCIN/AEI/10.13039/501100011033) and by the CERCA Programme / Generalitat de Catalunya. R.A. is supported by fellowship PRE2018-084278 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". We are grateful to Eva Borràs and Eduard Sabidó from the CRG/UPF Proteomics Unit for their advice and help in proteomics research.

5. References

1. Kaufmann K, Wellmer F, Muiño JM, Ferrier T, Wuest SE, Kumar V et al (2010)

Orchestration of floral initiation by APETALA1. Science 328(85):85–89.

2. Ó'Maoiléidigh DS, Thomson B, Raganelli A, Wuest SE, Ryan PT, Kwašniewska K et al (2015) Gene network analysis of Arabidopsis thaliana flower development through dynamic gene perturbations. Plant J 83(2):344–358.
3. Chen D, Yan W, Fu LY, Kaufmann K (2018) Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. Nat Commun 9:4534.
4. Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA et al (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol 15(R41).
5. Wuest SE, O'Maoileidigh DS, Rae L, Kwasniewska K, Raganelli A, Hanczaryk K et al (2012) Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. Proc Natl Acad Sci U S A 109(33):13452–13457.
6. Wellmer F, Riechmann JL (2010) Gene networks controlling the initiation of flower development. Trends Genet 26(12):519–527.
7. Wils CR, Kaufmann K (2017) Gene-regulatory networks controlling inflorescence and flower development in Arabidopsis thaliana. BBA - Gene Regul Mech 1860(1):95–105.
8. Pajoro A, Biewers S, Dougali E, Valentim FL, Mendes MA, Porri A et al (2014) The (r)evolution of gene regulatory networks controlling Arabidopsis plant reproduction: A two-decade history. J Exp Bot 65(17):4731–4745.
9. Heisler MG, Jönsson H, Wenkel S, Kaufmann K (2022) Context-specific functions of transcription factors controlling plant development: From leaves to flowers. Curr Opin Plant Biol 69:102262.

10. Takáč T, Šamajová O, Šamaj J (2017) Integrating cell biology and proteomic approaches in plants. J Proteomics 169:165–175.

11. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH (2017) Proteomics: Technologies and their applications. J Chromatogr Sci 55(2):182–196.

12. Grossmann J, Roschitzki B, Panse C, Fortes C, Barkow-Oesterreicher S, Rutishauser D et al (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 73(9):1740–1746.

13. Mergner J, Kuster B (2022) Plant Proteome Dynamics. Annu Rev Plant Biol 73:67–92.

14. Zhang Z, Hu M, Feng X, Gong A, Cheng L, Yuan H (2017) Proteomes and Phosphoproteomes of Anther and Pollen: Availability and Progress. Proteomics 17(20) doi: 10.1002/pmic.201600458.

15. Navrot N, Finnie C, Svensson B, Hägglund P (2011) Plant redox proteomics. J Proteomics 74(8):1450–1462.

16. Mergner J, Frejno M, List M, Papacek M, Chen X, Chaudhary A et al (2020) Mass-spectrometry-based draft of the Arabidopsis proteome. Nature 579(7799):409–414.

17. Kumar M, Carr P, Turner SR (2022) An atlas of Arabidopsis protein S-acylation reveals its widespread role in plant cell organization and function. Nat Plants 8(6):670–681.

18. Abraham P, Gannone RJ, Adams RM, Kalluri U, Tuskan GA, Hettich RL (2013) Putting the Pieces Together: High-performance LC-MS/MS Provides Network-, Pathway-, and Protein-level Perspectives in Populus. Mol Cell Proteomics 12(1):106-119.

19. Szymanski J, Levin Y, Savidor A, Breitel D, Chappell-Maor L, Heinig U et al (2017) Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. Plant J 90(2):396–417.

20. Duncan O, Trösch J, Fenske R, Taylor NL, Millar AH (2017). Resource: Mapping the Triticum aestivum proteome. Plant J 89(3):601–616.

21. Marx H, Minogue CE, Jayaraman D, Richards AL, Kwiecien NW, Siahpirani AF et al (2016) A proteomic atlas of the legume, M. truncatula, and its nitrogen fixing endosymbiont, S. meliloti. Nat Biotechnol 34(11):1198.

22. Dai S, Chen S (2012) Single-cell-type Proteomics: Toward a Holistic Understanding of Plant Function. Mol Cell Proteomics 11(12):1622-1630.

23. Emanuelsson O, Von Heijne G, Schneider G (2001) Analysis and prediction of mitochondrial targeting peptides. Methods Cell Biol 65:175–187.

24. Bruce BD (2000) Chloroplast transit peptides: Structure, function and evolution. Trends Cell Biol 10(10):440–447.

25. Bernhofer M, Goldberg T, Wolf S, Ahmed M, Zaugg J, Boden M et al (2018) NLSdb—major update for database of nuclear localization signals and nuclear export signals. Nucleic Acids Res 46(D1):D503–D508.

26. Bassal M, Abukhalaf M, Majovsky P, Thieme D, Herr T, Ayash M et al (2020) Reshaping of the Arabidopsis thaliana Proteome Landscape and Co-regulation of Proteins in Development and Immunity. Mol Plant 13(12):1709–1732.

27. Feng Z, Kong D, Kong Y, Zhang B, Yang X (2022) Coordination of root growth with root morphology, physiology and defense functions in response to root pruning in Platycladus orientalis. J Adv Res 36:187–199.

28. Jain A, Singh HB, Das S (2021) Deciphering plant-microbe crosstalk through proteomics studies. Microbiol Res 242:126590.

29. Niu Z, Liu L, Pu Y, Ma L, Wu J, Hu F et al (2021) iTRAQ-based quantitative proteome analysis insights into cold stress of Winter Rapeseed (Brassica rapa L.) grown in the field. Sci Reports 11:23434.

30. Koehler G, Rohloff J, Wilson RC, Kopka J, Erban A, Winge P et al (2015) Integrative "omic" analysis reveals distinctive cold responses in leaves and roots of strawberry, fragaria × ananassa 'Korona.' Front Plant Sci 6:826.

31. Le Signor C, Aimé D, Bordat A, Belghazi M, Labas V, Gouzy J et al (2017) Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. New Phytol 214(4):1597–1613.

32. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C et al (2018) Rewiring of the Fruit Metabolome in Tomato Breeding. Cell 172(1–2):249-261.e12.

33. Lehmann BD, Colaprico A, Silva TC, Chen J, An H, Ban Y et al (2021) Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. Nat Commun 12:6276.

34. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA et al (2018) Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. Brief Bioinform 19(2):286–302.

35. Kumar D, Bansal G, Narang A, Basak T, Abbas T, Dash D (2016) Integrating transcriptome and proteome profiling: Strategies and applications. Proteomics 16(19):2533–2544.

36. Wu Y, Tang Y, Jiang Y, Zhao D, Shang J, Tao J (2018) Combination of transcriptome sequencing and iTRAQ proteome reveals the molecular mechanisms determining petal shape in herbaceous peony (Paeonia lactiflora Pall.). Biosci Rep 38(6):BSR20181485.

37. Decourcelle M, Perez-Fons L, Baulande S, Steiger S, Couvelard L, Hem S et al (2015) Combined transcript, proteome, and metabolite analysis of transgenic maize seeds engineered for enhanced carotenoid synthesis reveals pleotropic effects in core metabolism. J Exp Bot 66(11):3141–3150.

38. Seaton DD, Graf A, Baerenfaller K, Stitt M, Millar AJ, Gruissem W (2018) Photoperiodic control of the Arabidopsis proteome reveals a translational coincidence mechanism. Mol Syst Biol 14(3):e7962.

39. Omidbakhshfard MA, Sokolowska EM, Di Vittori V, Perez de Souza L, Kuhalskaya A, Brotman Y et al (2021) Multi-omics analysis of early leaf development in Arabidopsis thaliana. Patterns 2(4):100235.

40. Huang Y, Zhou L, Hou C, Guo D (2022) The dynamic proteome in Arabidopsis thaliana early embryogenesis. Development 149(18):dev200715.

41. Keller M, Simm S, Bokszczanin KL, Bostan H, Bovy A, Chaturvedi P et al (2018) The coupling of transcriptome and proteome adaptation during development and heat stress response of tomato pollen. BMC Genomics 19(1):447

42. Ji J, Yang L, Fang Z, Zhuang M, Zhang Y, Lv H et al (2018) Complementary transcriptome and proteome profiling in cabbage buds of a recessive male sterile mutant provides new insights into male reproductive development. J Proteomics 179:80–91.

43. Xing M, Sun C, Li H, Hu S, Lei L, Kang J (2018) Integrated analysis of transcriptome and proteome changes related to the ogura cytoplasmic male sterility in cabbage. PLoS One 13(3):e0193462.

44. Jing D, Chen W, Hu R, Zhang Y, Xia Y, Wang S et al (2020) An integrative analysis of transcriptome, proteome and hormones reveals key differentially

expressed genes and metabolic pathways involved in flower development in loquat. Int J Mol Sci 21(14):5107.

45. Chen R, Chen G, Huang J (2017) Shot-gun proteome and transcriptome mapping of the jujube floral organ and identification of a pollen-specific S-locus F-box gene. PeerJ 5:e3588.

46. Lu D, Ni W, Stanley BA, Ma H (2016) Proteomics and transcriptomics analyses of Arabidopsis floral buds uncover important functions of ARABIDOPSIS SKP1-LIKE1. BMC Plant Biol 16:61.

47. Vogel C, Marcotte EM (2012) Insights into regulation of protein abundance from proteomics and transcriptomis analyses. Nat Rev Genet 13(4):227–232.

48. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4(10):1707–1719.

49. Taniguchi Y, Choi PJ, Li G, Chen H, Babu M, Hearn J et al (2011) Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science 329(5991):533–539.

50. Csárdi G, Franks A, Choi DS, Airoldi EM, Drummond DA (2015) Accounting for Experimental Noise Reveals that mRNA Levels, Amplified by Post-Transcriptional Processes, Largely Determine Steady-State Protein Levels in Yeast. PLoS Genet 11(5):1005206.

51. Oliva-Vilarnau N, Vorrink SU, Ingelman-Sundberg M, Lauschke VM (2020) A 3D Cell Culture Model Identifies Wnt/β-Catenin Mediated Inhibition of p53 as a Critical Step during Human Hepatocyte Regeneration. Adv Sci 7(15):2000248.

52. Simões T, Novais SC, Natal-da-Luz T, Devreese B, de Boer T, Roelofs D et al (2019). Using time-lapse omics correlations to integrate toxicological pathways of a formulated fungicide in a soil invertebrate. Environ Pollut 246:845–854.

53. Tarazona S, Balzano-Nogueira L, Conesa A (2018) Multiomics Data Integration in Time Series Experiments. Comprehensive Analytical Chemistry 82:505–532.

54. Lazar C, Gatto L, Ferro M, Bruley C, Burger T (2016) Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. J Proteome Res 15:1116–1125.

55. Jin L, Bi Y, Hu C, Qu J, Shen S, Wang X et al (2021) A comparative study of evaluating missing value imputation methods in label-free proteomics. Sci Rep 11:1720.

56. Martínez-García JF, Monte E, Quail PH (1999) A simple, rapid and quantitative method for preparing Arabidopsis protein extracts for immunoblot analysis. Plant J 20:251–257.

57. Beer LA, Liu P, Ky B, Barnhart KT, Speicher DW (2017) Efficient quantitative comparisons of plasma proteomes using label-free analysis with MaxQuant. Methods Mol Biol 1619:339–352.

58. Gerster S, Kwon T, Ludwig C, Matondo M, Vogel C, Marcotte EM et al (2014) Statistical approach to protein quantification. Mol Cell Proteomics 13(2):666–677.

59. Fabre B, Lambour T, Bouyssié D, Menneteau T, Monsarrat B, Burlet-Schiltz O et al (2014) Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. EuPA Open Proteomics 4:82–86.

60. Meleth S, Deshane J, Kim H (2005) The case for well-conducted experiments to validate statistical protocols for 2D gels: Different pre-processing = different lists of significant proteins. BMC Biotechnol 5:7.

61. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R et al (2001) Missing value estimation methods for DNA microarrays. Bioinformatics

17(6):520–525.

62. Wellmer F, Alves-Ferreira M, Dubois A, Riechmann L, Meyerowitz EM (2006) Genome-Wide Analysis of Gene Expression during Early Arabidopsis Flower Development. PLoS Genet 2(7):e117.

63. Eng JK, Mccormack AL, Yates JR (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. Am Soc Mass Spectrom 5:977–989.

64. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probablity-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3557.

65. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. Proteomics 3(8):1454–1463.

66. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra.cBioinformatics 20(9):1466–1467.

67. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM et al (2004) Open Mass Spectrometry Search Algorithm. J Proteome Res 3:958–964.

68. Fu Y, Yang Q, Sun R, Li D, Zeng R, Ling CX et al (2004) Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. Bioinformatics 20(12):1948–1954.

69. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M et al (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem 77(14):4626–4639.

70. Bern M, Cai Y, Goldberg D (2007) Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass

Spectrometry. Anal Chem 79(4):1393–1400.

71. Eng JK, Jahan TA, Hoopmann MR (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics 13(1):22–24.

72. Kim S, Pevzner PA (2014) MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun 5(1):5277.

73. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11(12):2301–2319.

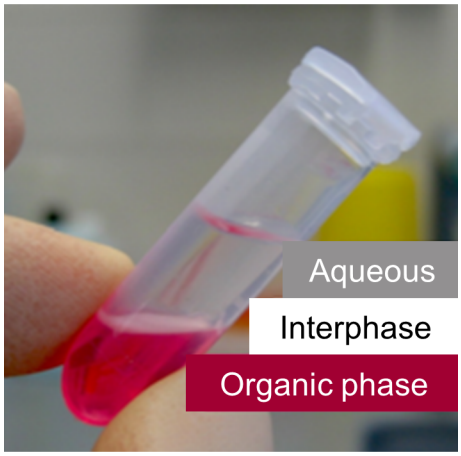
74. Zeng X, Ma B (2021) MSTracer: A Machine Learning Software Tool for Peptide Feature Detection from Liquid Chromatography-Mass Spectrometry Data. J Proteome Res 20(7):3455–3462.

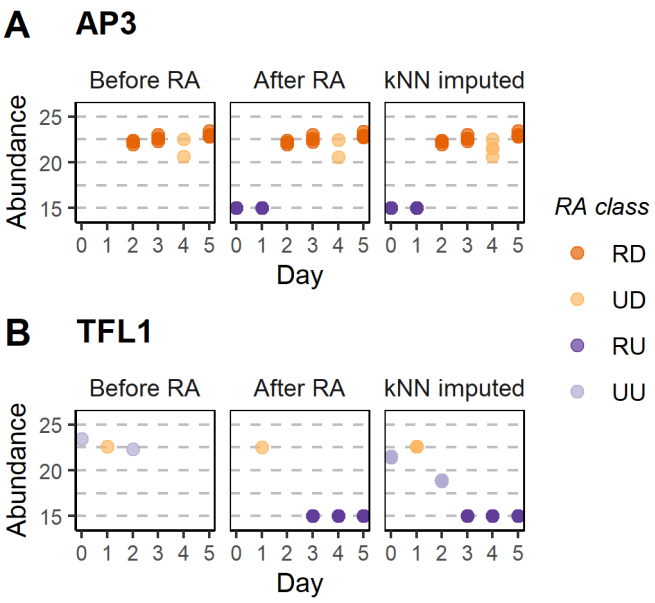
75. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 50(D1):D543–552.

Figure Captions

Figure 1. Picture of the three phases formed in step 4 of the protein extraction method (section 3.1).

Figure 2. Stringent analysis to identify reliably undetected and detected fraction of a proteome. The analysis allows to impute values for MAR and MNAR considering their biological meaning. The figure illustrates results from a time course experiment using the Arabidopsis floral induction system pAP1:AP1-GR *ap1cal* ***(1)***, in which samples were collected at one-day intervals after floral induction (day 0), up to day 5. Log2 TOP3 abundances through time of two flower development regulators, APETALA 3 (AP3) (A) and TERMINAL FLOWER 1 (TFL1) (B), before and after the "reliability analysis" (RA), and after kNN imputation (from left to right) (n = 4 biological replicates).


Aqueous
Interphase
Organic phase


A
AP3
Before RA
After RA
kNN imputed
Abundance
25
20
15
0 1 2 3 4 5
Day
B
TFL1
RA class
RD
UD
RU
UU