From single cells to flowers – Biological complexity driving plant reproductive development

Isabel Schwarz[1]
ias@plen.ku.dk - ORCID: 0000-0003-4717-422X

Manuel Neumann[2]
manuel.neumann@hu-berlin.de - ORCID: 0000-0002-7883-0843

Rosario Vega[2]
rosario.vega@hu-berlin.de - ORCID: 0000-0002-7104-8111

Xiaocai Xu[2]
xiaocai.xu@hu-berlin.de

Letizia Cornaro[3]
letizia.cornaro@unimi.it – ORCID: 0000-0001-8488-6731

Lucia Colombo[3]
lucia.colombo@unimi.it – ORCID: 0000-0001-8415-1399

Teva Vernoux[4]
teva.vernoux@ens-lyon.fr - ORCID: 0000-0002-8257-4088

Jian Xu[5]
J.Xu@science.ru.nl – ORCID: 0000-0001-6194-6316

Sebastian Marquardt[1, *]
sebastian.marquardt@plen.ku.dk – ORCID: 0000-0003-1709-2717

Kerstin Kaufmann[2, *]
kerstin.kaufmann@hu-berlin.de - ORCID: 0000-0001-7960-6256

Word count: 1531

[1] Department of Plant and Environmental Sciences, Copenhagen Plant Science Centre, Københavns Universitet, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

[2] Department for Plant Cell and Molecular Biology, Institute for Biology, Humboldt-Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany

[3] Dipartimento di Bioscienze, Università degli Studi di Milano, Via Giovanni Celoria 26, 20133 Milano, Italy

[4] Laboratoire Reproduction et Développement des Plantes, Univ Lyon, ENS de Lyon, CNRS, INRAE F-69342, Lyon, France

[5] Department of Plant Systems Physiology, Institute for Water and Wetland Research, Radboud University, Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands

*Co-Corresponding authors: Sebastian Marquardt and Kerstin Kaufmann

----

ABSTRACT

The rise of data science in biology stimulates interdisciplinary collaborations to address fundamental questions. Here, we report the outcome of the first SINFONIA symposium focused on revealing the mechanisms governing plant reproductive development across biological scales. The intricate and dynamic target networks of known regulators of flower development remain poorly understood. To analyze development from the genome to the final floral organ morphology, high-resolution data that capture spatiotemporal regulatory activities are necessary and require advanced computational methods for analysis and modeling. Moreover, frameworks to share data, practices and approaches that facilitate the combination of varied expertise to advance the field are called for. Training young researchers in interdisciplinary approaches and science communication offers the opportunity to establish a collaborative mindset to shape future research.

KEYWORDS

flower development; epigenetics; RNA biology; Genomics; single cell biology

TEXTBOX
From single cells to flowers – transcriptional and post-transcriptional mechanisms driving plant development with the acronym SINFONIA represents a bottom-up initiative of research groups from 5 countries. Each partner performs research on the regulation and coordination of plant genomes. SINFONIA promises an outlet to coordinate individual efforts to unlock synergies in the area of reproductive plant biology, specifically flower development. The initiative is currently unfunded; the kick-off meeting of SINFONIA was a community organized online symposium. SINFONIA is planning future symposia, networking platforms and training activities with the vision that these will be open to interested parties in the community.

Join SINFONIA
We welcome other groups to participate in SINFONIA. The best way to join SINFONIA is by registering for our mailing list or following us on twitter < twitter handle >. These channels will be an initial stepping stone to hear about updates on SINFONIA partners regarding research efforts, possibilities of collaboration, job openings and events (e.g. PhD training) organized by SINFONIA.

From single cells to flowers – Biological complexity driving plant reproductive development

Introduction

The rise of data science in biology stimulates interdisciplinary collaborations to address fundamental questions. Here, we report the outcome of the first SINFONIA symposium focused on mechanisms governing plant reproductive development across biological scales. To analyze development from genome to floral organ morphology, high-resolution data that capture spatiotemporal regulatory activities are necessary and require advanced computational methods for analysis and modeling. Frameworks to share data and protocols that facilitate the combination of varied expertise are called for. Training young researchers in interdisciplinary approaches and science communication offers the opportunity to establish a collaborative mindset shaping future research.

Open questions on the regulatory basis of flower formation

To better understand the processes underlying flower formation, we need to consider a range of key molecular events across time, space and scales. While key signals and genes underlying flower development have been identified, the precise transcriptional and post-transcriptional mechanisms that direct the spatiotemporal diversity in development are less clear.

In the classical model of transcriptional regulation, the presence of cis-regulatory DNA elements and the combination of a specific pool of transcription factors (TFs) and chromatin regulators

(Kaufmann *et al.*, 2010) regulate gene expression at the level of transcriptional initiation. Here, the spatiotemporal heterogeneity of TF abundance controls how cells with distinct functions and behaviors are specified. This process is fine-tuned by TFs that act in concert with various species of non-coding RNAs (ncRNAs) (Chen *et al.*, 2020; Samad *et al.*, 2017; Ivanov *et al.*, 2021). Once initiated, gene expression can further be controlled by mechanisms such as promoter-proximal stalling at the level of transcriptional elongation (Thomas *et al.*, 2020; Leng *et al.*, 2020; Kindgren *et al.*, 2020). While one can build strong hypotheses for the binding of TFs to DNA that likely trigger transcriptional activation of genes, identification and functional prediction of ncRNA and elongation-based control mechanisms remain a key challenge.

Since factors controlling transcriptional regulation depend on the chromatin state, which is defined by chromatin architecture, histone modifications, DNA-methylation and DNA accessibility, epigenetics offers another layer of gene regulation (Leng *et al.*, 2020). Understanding the relation between the chromatin landscape and TFs promises insight into the question of how one genome spawns the different cell types and cell states. This is well in line with the incentives of the Plant Cell Atlas initiative (Rhee *et al.*, 2019), and aims to generate opportunities for advancing our knowledge on plant reproductive development. This relationship is particularly important in processes like flowering time control where environmental sensing is crucial for a correct progression (Whittaker & Dean, 2017).

Finally, post-transcriptional events regulating RNA structure, mRNA transcript stability, protein–RNA interactions and ribosome occupancy offer an interface to adjust the translation machinery and, in consequence, the protein levels (Merchante *et al.*, 2017), which decide on a cell's function within the developing flower meristem.

We envision that a detailed conceptual understanding of how these processes control the progression of flower development will depend on the spatial context of cells. This is particularly important in the context of flower development because a) flower tissues are not as well defined as e.g. roots, b) cell fates are harder to predict due to the absence of stereotypical cell division patterns, c) stem cells in the floral meristem seemingly lack the level of organization present in the root stem cell niche and d) the positional information of cells influences their lineage-commitment.

Technical challenges in resolving the complexity of flower development

The elucidation of the regulatory processes described here calls for the establishment of new experimental and computational methodologies that describe cellular states at a higher resolution in space and time.

A promising development in this direction is given by the emergence of single-cell omics technologies which offer a great opportunity to refine our understanding of the different layers of gene regulation by enabling us to discern cellular heterogeneity within tissues and organs (Ryu *et al.*, 2019; Shaw *et al.*, 2021; Engelhorn *et al.*, 2014; You *et al.*, 2017; Rahmani *et al.*, 2019). Despite the ability to be able to collect information about the genome, epigenome, and proteome (Fang *et al.*, 2019; Kelsey *et al.*, 2017; Marx *et al.*, 2019) at the single cell level, one key challenge is to apply those technologies to the flower.

Besides a higher resolved description of gene regulatory processes, the spatial position of cells in the developing flower meristem also has to be considered. Inconveniently, traditional methods for

DNA or RNA extraction require freezing and disruption of tissues, that results in the loss of the positional information of cells. An approach to preserve spatial information could combine sequencing with single molecule imaging using super-resolution microscopy (Halpern *et al*., 2017). Alternatively, computational methods allow the *de novo* reconstruction of single-cell spatial gene expression from scRNA-seq data *in silico* (Nitzan *et al.,* 2019). The 3D reconstruction of gene activities (Neumann *et al.,* 2021, Preprint) in tissues described here currently represents a bottleneck to thoroughly understand the trajectory from molecular signals to flower organs, and requires advanced imaging and computational methodologies for tissue reconstruction (Vijayan *et al.,* 2021). The establishment of such tools together with state of the art spatial-genomics sequencing technologies will pave the way towards a comprehensive genome-wide 3D expression atlas in plants.

Next to the spatially resolved description of cells, the study of developmental processes requires the incorporation of a time axis. This can be achieved by collecting plant tissue at different developmental stages followed by sequencing. The disconnected datasets can be subjected to pseudo-time analysis that groups cells in a lower dimensional space so that lineages and developmental trajectories become apparent. Alternatively, live imaging can capture the continuous changes of cellular components during development directly. Although informative, this approach suffers from the challenge of keeping plant tissue at low levels of stress during long-time observations and prevents the scaling to high-throughput approaches.

Despite the potential of spatio-temporally resolved single-cell data from different modalities to improve our understanding of the processes shaping flower development, a key challenge is to harness the information contained within that data to its fullest potential. This requires the integration of multi-modal data and correction for batch-effects that arise for example due to the application of different sequencing technologies. (Butler *et al.*, 2018; Stuart *et al.*, 2019)

Lastly, the integrated data can be used to construct gene regulatory networks, integrating predicted cis-regulatory elements (Li *et al.*, 2018) and their logic (Cofer *et al.*, 2020, Preprint). Endeavors in this direction are expanding current approaches towards understanding the causality of gene expression from single-cell data (Yuan et *al.*, 2019) and may allow phenotype prediction based on a given DNA sequence (Zhou *et al.*, 2018).

Synergistic research strategies

Developing new experimental and computational methods to address prevalent biological questions requires connecting international groups with expertise across biological levels and experimental procedures. Using an interdisciplinary and integrative approach to address one biological problem will empower SINFONIA researchers to jointly unravel the mechanisms that dictate flower development across biological scales and regulatory levels. Collaborative efforts will allow to scale new-generation high throughput techniques to investigate gene expression, regulatory processes, and translational efficiency down to the single-cell level.

Sharing of expertise and data about these developments is essential to maximize research output. In this context, collaboration can be enhanced through IT infrastructure such as an online platform that contains e.g. information about experimental procedures, deposits analyzed results and enables easy communication. The combination of these functions into one platform will enable SINFONIA researchers to learn about and minimize the effort required to implement cutting edge

protocols, to obtain help from experts almost instantaneously and ensure that research efforts of individual groups are designed towards reaching the overarching goal of SINFONIA.

In addition, bioinformatics experts will be essential to aid the analysis of complex multi-modal single-cell data, develop novel computational approaches and provide the community with an interactive 4D plant model that combines results from different experiments to facilitate accessibility of research data and key concepts to a broad audience.

In order to foster the interdisciplinary character of SINFONIA, those efforts can be complemented with on-site training, knowledge dissemination about novel methodologies as well as the shared usage of experimental infrastructure.

Interdisciplinary PhD training

Integration of diverse knowledge is a significant benefit of interdisciplinary research but also poses substantial challenges to PhD students and consortium members. Workshops teaching the vocabulary and methodology of e.g. computational scientists to experimental biologists and *vice versa* can contribute to overcome communication barriers between scientists of diverse fields and lead to a better overall research design focused on flower development. This research brings the classical field of plant cell biology to a new level, raising the need for integrated understanding of molecular regulatory and physiological processes with tissue growth and mechanics. Essential for this integration and the success of visionary research is the improvement of communication skills through schooling of presentation techniques, public speaking to scientific and general audiences and writing of publications and grant applications. It should be a goal to make science attractive and easily comprehensible.

Naturally, all participants need to share an understanding of the overarching goal of the research and of the current state of knowledge within the consortium. Presenting recent progress in individual projects at regular symposia fosters connection between colleagues, the identification of knowledge gaps and more targeted collaborations. Close association of groups with various specialties enables support for Ph.D. students by diverse supervisory committees to aid and train them in approaching questions multi-dimensionally.

Conclusion

In summary, SINFONIA aims at improving our understanding of flower development by developing novel single-cell technologies that allow the reconstruction of spatiotemporal dynamics of cell identities across biological scales.

Clearly, there is a great need to improve experimental and computational methods to unlock the benefits of emerging single cell technologies for plant science. The increasing volume and complexity of spatiotemporally resolved data requires international collaborations of specialized groups. Dedicated IT infrastructure is becoming essential for collaboration, and young researchers require specialized training to foster interdisciplinary perspectives. Those represent the key goals of the SINFONIA initiative.

Acknowledgements

Key words

big data, data science, flower development, gene regulation, interdisciplinary research, machine learning, PhD training, single-cell OMICs, systems biology, transcription

Author contribution statement

I.S., M.N., R.V., X.X. and L.C. did writing in the original draft, as well as review and editing. T.V., L.C., J.X., S.M. and K.K. conceived the topic and edited the manuscript.

References

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology 36(5), 411-420.

Chen L, Zhu Q-H, Kaufmann K. 2020. Long non-coding RNAs in plants: Emerging modulators of gene activity in development and stress responses. Planta 252(5), 92.

Cofer EM, Raimundo J, Tadych A, Yamazaki Y, Wong AK, Theesfeld CL, Levine MS, Troyanskaya OG. 2020. DeepArk: modeling cis-regulatory codes of model species with deep learning. BioRxiv https://doi.org/10.1101/2020.04.23.058040. [Preprint].

Engelhorn, J.; Blanvillain, R.; Carles, C.C. Gene activation and cell fate control in plants: A chromatin perspective. Cell. Mol. Life Sci.2014,71, 3119–3137.

Fang Z, Weng C, Li H, *et al*. 2019. Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key β-Cell-Specific Disease Genes. Cell Reports 26(11), 3132-3144.e7.

Halpern KB, Shenhav R, Matcovitch-Natan O, *et al*. 2017. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542(7641), 352-356. Erratum in: Nature. 2017. 543(7647), 742.

Ivanov M, Sandelin A, Marquardt S. 2021. TrancriptomeReconstructoR: data-driven annotation of complex transcriptomes. BMC Bioinformatics. 22(1),290. Erratum in: BMC Bioinformatics. 2021. 22(1), 370.

Kaufmann K, Pajoro A, Angenent GC. 2010. Regulation of transcription in plants: Mechanisms controlling developmental switches. Nature Reviews Genetics 11(12), 830–842.

Kelsey G, Stegle O, Reik W. 2017. Single-cell epigenomics: Recording the past and predicting the future. Science 358(6359), 69-75.

Kindgren P, Ivanov M, Marquardt S. 2020. Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in Arabidopsis. Nucleic Acids Res. 48(5), 2332-2347.

Leng X, Thomas Q, Rasmussen SH, Marquardt S. 2020. A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. Trends Plant Sci. 25(8), 744-764.

Leng X, Ivanov M, Kindgren P, Malik I, Thieffry A, Brodersen P, Sandelin A, Kaplan CD, Marquardt S. 2020. Organismal benefits of transcription speed control at gene boundaries. EMBO Rep. 21(4),e49315.

Li Y, Shi W, Wasserman WW. 2018. Genome-wide prediction of cis-regulatory regions using supervised deep learning methods. BMC Bioinformatics 19(1), 202.

Marx V. 2019. A dream of single-cell proteomics. Nature Methods 16(9), 809-812.

Merchante C, Stepanova A, Alonso J. 2017. Translation regulation in plants: an interesting past, an exciting present and a promising future. The Plant Journal 90(4), 628-653.

Neumann, M., Xu, X., Smaczniak, C., Schumacher, J., Yan, W., Blüthgen, N., Greb, T., Jönsson, H., Traas, J., Kaufmann, K., & Muino, J. M. (2021). A 3D gene expression atlas of the floral meristem based on spatial reconstruction of single nucleus RNA sequencing data. BioRxiv, 2021.06.30.450319. [Preprint].

Nitzan M, Karaiskos N, Friedman N, Rajewsky N. 2019. Gene expression cartography. Nature 576(7785), 132-137.

Rahmani, E., Schweiger, R., Rhead, B. *et al.* Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. Nat Commun 10, 3417 (2019).

Rhee SY, Birnbaum KD, Ehrhardt DW. Towards Building a Plant Cell Atlas. Trends Plant Sci. 2019;24(4):303-310.

Ryu KH, Huang L, Kang HM, Schiefelbein J. 2019. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant Physiology 179(4), 1444-1456.

Shaw R, Tian X, Xu J. 2021. Single-Cell Transcriptome Analysis in Plants: Advances and Challenges. Molecular Plant 14(1), 115-126.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell, 177(7).

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I. 2017. MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. Frontiers in Plant Science 8, 565.

Thomas QA, Ard R, Liu J, Li B, Wang J, Pelechano V, Marquardt S. 2020. Transcript isoform sequencing reveals widespread promoter-proximal transcriptional termination in Arabidopsis. Nature Communications 11(1), 2589.

Whittaker C, Dean C. 2017. The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. Annual Review of Cell and Developmental Biology 33(1), 555–575.

You, Y., Sawikowska, A., Neumann, M. *et al.* Temporal dynamics of gene expression and histone marks at the Arabidopsis shoot meristem during flowering. Nat Commun 8, 15120 (2017).

Yuan Y, Bar-Joseph Z. 2019. Deep learning for inferring gene relationships from single-cell expression data. Proceedings of the National Academy of Sciences of the USA 116(52), 27151–27158.

Vijayan A, Tofanelli R, Strauss S, Cerrone L, Wolny A, Strohmeier J, Kreshuk A, Hamprecht FA, Smith RA, Schneitz K. 2021. A digital 3D reference atlas reveals cellular growth patterns shaping the Arabidopsis ovule. eLife 10:e63262.

Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. 2018. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nature Genetics 50(8),1171-1179.