SHORT COMMUNICATION

Integrating Single Nuclei and Bulk RNA Sequencing in Rice Shoot Apical Meristems Uncovers Candidate Early Floral Transition Gene Networks

Daniele Traversa[1]*, Giulio Vicentini[2]*, Paolo Korwin Krukowski[1]*, Lucio Conti[1], Matteo Chiara**[1] and Vittoria Brambilla[2]**

1) Department of Biosciences, Università degli Studi di Milano, via Celoria 26, 20133 Milan, Italy
2) Department of Agricultural and Environmental Sciences, Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy

*these authors contributed equally
**corresponding authors

VB, LC, MC designed experiments; GV, PKK, DT performed experiments; DT, GV, PKK, LC, MC, VB wrote manuscript

ABSTRACT

In rice, short days trigger floral transition and the transcriptional reprogramming of the shoot apex to become reproductive. We integrated time-resolved bulk RNA-seq with single nuclei RNA-seq analysis to gain a refined understanding of the transcriptional programs induced at the shoot apex during floral transition. Our analyses highlighted technological and conceptual differences between single nuclei RNA-seq and bulk RNA-seq data and described previously uncharacterized transcriptional programs associated with the early steps of floral induction in rice.

MAIN

In rice, short-day (SD) photoperiods trigger the expression of the florigen, a systemic signal produced in leaves that can activate floral transition at the shoot apex. Florigen(s) are encoded by the *Hd3a* and *RFT1* genes and move within the phloematic stream to the shoot apical meristem (SAM) to cause reprogramming of meristematic cells from a vegetative to an

inflorescence meristem identity (Brambilla and Fornara 2013). Long day (LD)-SD-LD double shift experiments (Gómez-Ariza et al. 2019) showed that the commitment of the SAM to an inflorescence identity requires 12 consecutive SDs, after which reversion to the vegetative state is no longer possible. Defining the earliest transcriptional changes at the SAM is key to understanding how cell fates are reprogrammed and SAM commitment is established. Time-resolved bulk RNA-seq of SAMs undergoing floral transition, a commonly used approach to capture the transcriptome from entire meristematic tissues, may overlook local expression patterns due to the high cellular heterogeneity of the samples. A breakthrough technology, single cell/single nuclei sequencing, could overcome some of these limitations. To gain a refined understanding of the transcriptional programs induced at the SAM in specific cell types, we exposed rice plants to an increasing number of SDs and harvested SAMs for single nuclei RNA-seq (snRNA-seq). Six-week-old LD-grown plants were exposed to 0 (TP1), 5 (TP2), 9 (TP3) and 13 (TP4) SDs and morphological changes of the SAM were analyzed under a stereomicroscope (Figure 1A). At every time point SAMs- containing about 200 - 250 cells at these stages (Nosaka-Takahashi et al. 2022) were harvested from 8 independent biological replicates, manually dissected and pooled. Nuclei were sorted by flow cytometry, and a total of 7000 nuclei were loaded for snRNA-seq analysis.

Following barcode deconvolution (see Supplementary Material and Methods) 6116 distinct nuclei were detected, and of these 5938 passed quality filters. An average of 1485 nuclei per time point were analyzed (min:1140, max:1901). An average number of 646 UMI per nucleus (min 217; max 7011; Supplementary Fig. S1) was recorded, while the mean value of detected transcripts per nucleus was 493 (min 200; max 3673; Supplementary Fig. S2). Descriptive read quality statistics and the proportion of mapped reads are summarized in Supplementary Table S1. Data were pooled across time points and analyzed as a continuum. The standard Seurat workflow partitioned the 5938 nuclei into 11 distinct clusters (Supplementary Fig. S3 and Supplementary Table S2). Every cluster included a comparable number of nuclei (avg 545; min 124; max 923) and visual inspection of the UMAP plot did not suggest neat separations between clusters. This pattern is compatible with a continuous differentiation of transcriptional programs at the early stages of floral induction at the SAM and aligns with previous findings (Satterlee et al 2020). Although no neat one-to-one correspondence between clusters and time points was observed, cluster 2 was strongly associated with time point 2, cluster 10 with time point 3 and cluster 11 with time point 4 (Supplementary Fig. S4). A total of 3678 genes were specifically associated with one or more of the 11 clusters defined by Seurat (Supplementary Table S3). Functional enrichment analyses of biological process gene ontology (GO) terms identified both common and specific patterns of enrichment. GO terms associated with “membrane”, “RNA binding” and “protein binding” were enriched throughout all clusters (albeit at different levels) (Figure 1B). Cluster 7 exhibited a significant

enrichment in: “microtubule binding” and “microtubule motor activity”. Cluster 10 was enriched in terms associated with “DNA replication” and “chromatin organization” and Cluster 11 showed significant enrichment in “protein transport”, and “lipid metabolic processes”. Semantic similarity analyses of enriched ontologies (Cao et al. 2023) defined 6 main topics (Topic 1-6) of functional enrichment (Supplementary Fig. S5). Terms associated with transcriptional regulation/DNA binding were shared by Topic 1, Topic 2 and Topic 6 and associated with several clusters (Cluster 1, Cluster 2, Cluster 4, Cluster 5, Cluster 10 and Cluster 11), conversely, some topics and functional enrichments were associated specifically to a cluster. For example, Topic 3 corresponded with Cluster 8 and terms associated with plasma membranes; Topic 6 was associated with Cluster 10 and strongly enriched in terms associated with chromatin remodeling and organization. Pseudotime trajectory inference (Cao et al. 2019) identified 9 distinct programs of gene expression (Supplementary Table S4); an almost complete one-to-one correspondence with the 11 clusters defined by Seurat (Supplementary Fig. S6-8) was observed and genes associated with distinct pseudotime modules largely overlapped with cluster differentially expressed genes (modules: tot. 4495 DEGs, clusters tot. 3678 DEGs, common 2933 DEGs). Interestingly, Cluster 1, Cluster 5 and Cluster 2, which were enriched in genes associated with transcriptional regulation according to our functional enrichment analyses, were mapped at the early stages of the pseudotime progression. Cluster 7 and Cluster 10, which are enriched in genes involved in microtubule binding and chromatin remodeling, were aligned with a later stage of development. These observations are not incompatible with early transcriptional reprogramming, followed by an increased rate of DNA duplication and cell growth at later stages of development.

A manually curated short list of markers with well-characterized spatial expression at the SAM (i.e. by *in situ* hybridization), was used to attribute cellular states to the clusters identified by Seurat (Supplementary Table S5, *in situ* genes). Unfortunately, most of these genes were not found to be strongly associated with any of our clusters (Supplementary Fig. 9A) and were detected in a limited number of nuclei (85% of genes in less than 10% of nuclei Supplementary Fig. S10A). Similarly, floral transition upregulated genes, as identified by differential gene expression analyses in a bulk RNA-seq experiment (Supplementary Table 5, bulk-RNAseq Mineri et al. 2023), were barely detectable in our experimental settings (Supplementary Fig. S9B) and were expressed in less than 10% of the nuclei (Supplementary Fig. S10B). This notwithstanding, when snRNA-seq gene expression was aggregated *in-silico* by time point (pseudo-bulk analysis) and compared with a bulk RNA-seq with a similar experimental layout (Gómez-Ariza et al. 2019). A statistically significant correlation was recovered between gene expression levels of the 23312 genes commonly expressed in both datasets, at every time point (Supplementary Fig. S11A) and irrespective of gene expression levels as estimated from the bulk RNA-seq data (Supplementary Fig. S11B). Likewise, the correlation of expression

levels between marker/selected genes included in Supplementary Table 5 across the 2 datasets (Supplementary Fig. 12) was also statistically significant.

Cluster-specific genes identified by our analyses were compared with markers of meristem-associated cellular identities/states as identified by Zong et al. at by snRNA-seq at a later stage of flowering in rice (Zong et al. 2022). Clusters 7 and 8 specific genes had a significant overlap with “spikelet meristem”; Cluster 4 with “branch meristem” and Clusters 5 and 3 with “inflorescence meristem” (Supplementary Fig. S13).

In-silico analyses of promoter sequences were performed to reconcile transcriptional programs as inferred from differentially expressed genes identified by Gómez-Ariza et al. 2019 at different time points, genes differentially expressed in distinct clusters and genes differentially expressed between different pseudotime modules. The pscan (Zambelli et al. 2009) software and the complete non-redundant collection of plant transcription factor binding sites (TFBS) as provided by the Jaspar database (Sandelin et al. 2004), were used to identify statistically significant enriched families of transcription factors. Three groups of TFBS enrichment profiles were delineated (Figure 1C). Group 1 and Group 2 included most clusters/modules and time point 2 (4 SD) from the bulk RNA-seq dataset, whereas Group 3 included only Cluster 9 and Module 5 and was loosely associated with time point 4 (12 SD) from Gómez-Ariza et al. 2019. Both Group 1 and Group 2 were enriched in transcription factors belonging to the TCP and FRS/FRF family/class (Figure 1D); MYB-related TFs were over-represented exclusively in Group 1 while several classes/families were specifically enriched only in Group 2. Group 3 was significantly enriched in TFs from the DOF and C2H2 families and a significant enrichment of the C2H2 family was observed also when a larger group (see dotted line in Figure 1C), formed by Module 5 and 8, Cluster 9 and Timepoint 12 SD of the bulk RNA-seq was considered.

Manual analyses of expression profiles aggregated by time point (pseudo-bulk analysis) identified 88 genes with a large change in expression ($\log2(FC) \geq 2$ or $\log2(FC) \leq -2$ in at least one time point (Supplementary Table S6) and associated with the previously identified significantly over-represented transcription factor families. Interestingly, GO functional enrichment analyses according to the Panther database (Mi et al. 2019) highlighted a statistically significant enrichment of GO terms associated with flowering, including regulation of secondary shoot formation (GO:2000032), regulation of plant organ formation (GO:1905428); ethylene-

activated signaling pathway (GO:0009873); cellular response to ethylene stimulus (GO:0071369) and response to ethylene (GO:0009723) for these genes.

In conclusion, in our experimental condition, snRNA-seq provided complementary results to bulk RNA-seq but also demonstrated some limitations. Since the relatively low number of UMI/unique molecules counted per nucleus, contemporary snRNA-seq approaches offer only a partial representation of the transcriptome and might sometimes fail to recapitulate patterns of gene expression determined by analytical approaches with a higher resolution (*in situ* hybridization) and/or a broader scope (bulk RNA-seq). In our opinion, these considerations, coupled with the lack of a large body of data and/or curated databases of cellular identities, such as those available for *H.sapiens* (Tabula Sapiens Consortium et al. 2022), are currently the main limiting factor for the systematic application of snRNA-seq in crops. While approaches for mapping cellular identities in plants are becoming increasingly effective (Nobori et al. 2023), unraveling the transcriptional dynamics and cellular identities within the SAM remains challenging (Satterlee et al. 2020). These challenges arise from the difficulty in accessing the SAM with minimal manipulation, its intricate cellular organization, and the limited number of cells comprising this organ.

Despite these limitations, we observe that our results recapitulate previous findings (Zong et al. 2022, Satterlee et al. 2020) and align with transcriptional dynamics as determined by bulk RNA-seq under similar conditions (Gómez-Ariza et al. 2019). Moreover, by performing a systematic analysis of transcription factor binding sites we could capture a fine grained snapshot of transcriptional programs and transcription factor families associated with floral transition at specific time points. Our analyses identified 88 TFs with large changes in expression and belonging to TF families that regulate a large proportion of our cluster-specific genes. These TFs might represent a previously uncharacterized ensemble of early regulators of floral induction. Remarkably, a functional enrichment gene ontology analysis suggests a strong association with flowering related developmental processes (Supplementary Table 6) for some of these TFs. Although our results uncovered previously uncharacterized transcriptional programs associated with the early steps of floral induction in rice, further validation through spatial transcriptomics might be necessary to reconstruct spatially resolved patterns of expression of these TFs.

REFERENCES

Brambilla V, Fornara F (2013) Molecular Control of Flowering in Response to Day Length in Rice. J Integr Plant Biol 55:. https://doi.org/10.1111/jipb.12033

Cao J, Spielmann M, Qiu X, et al (2019) The single-cell transcriptional landscape of mammalian organogenesis. Nature 566:. https://doi.org/10.1038/s41586-019-0969-x

Cao S, He Z, Chen R, et al (2023) scPlant: A versatile framework for single-cell transcriptomic data analysis in plants. Plant Commun 4:. https://doi.org/10.1016/j.xplc.2023.100631

Gómez-Ariza J, Brambilla V, Vicentini G, et al (2019) A transcription factor coordinating internode elongation and photoperiodic signals in rice. Nat Plants 5:. https://doi.org/10.1038/s41477-019-0401-4

Li C, Zhang S, Yan X, et al (2023) Single-nucleus sequencing deciphers developmental trajectories in rice pistils. Dev Cell 58:. https://doi.org/10.1016/j.devcel.2023.03.004

Mi H, Muruganujan A, Huang X, et al (2019) Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc 14:. https://doi.org/10.1038/s41596-019-0128-8

Mineri L, Cerise M, Giaume F, et al (2023) Rice florigens control a common set of genes at the shoot apical meristem including the F-BOX BROADER TILLER ANGLE 1 that regulates tiller angle and spikelet development. Plant Journal 115:1647–1660. https://doi.org/10.1111/tpj.16345

Nobori, T., Oliva, M., Lister, R. et al. (2023) Multiplexed single-cell 3D spatial gene expression analysis in plant tissue using PHYTOMap. Nat. Plants 9, 1026–1033. https://doi.org/10.1038/s41477-023-01439-4

Nosaka-Takahashi M, Kato M, Kumamaru T, Sato Y (2022) Measurements of the number of specified and unspecified cells in the shoot apical meristem during a plastochron in rice (Oryza sativa) reveal the robustness of cellular specification process in plant development. PLoS One 17:. https://doi.org/10.1371/journal.pone.0269374

Sandelin A, Alkema W, Engström P, et al (2004) JASPAR: An open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:. https://doi.org/10.1093/nar/gkh012

Satterlee JW, Strable J, Scanlon MJ. Plant stem-cell organization and differentiation at single-cell resolution. Proc Natl Acad Sci U S A. 2020 Dec 29;117(52):33689-33699. doi: 10.1073/pnas.2018788117. Epub 2020 Dec 14. PMID: 33318187; PMCID: PMC7776995.

Tabula Sapiens Consortium*, Jones RC, Karkanias J, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022;376(6594):eabl4896. doi:10.1126/science.abl4896

Zambelli F, Pesole G, Pavesi G (2009) Pscan: Finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res 37:. https://doi.org/10.1093/nar/gkp464

Zong J, Wang L, Zhu L, et al (2022) A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. New Phytologist 234:. https://doi.org/10.1111/nph.18008

Acknowledgements

Fundings for this project derive from the “Bando Straordinario per Progetti Interdipartimentali (SEED)” of the University of Milan 2019 – Deciphering light stimulated transcriptional reprogramming in plants at single cell resolution- DISENGAGE obtained by Vittoria Brambilla, Lucio Conti and Matteo Chiara. We thank Dr. Claudia Bazzini (Flow cytometry facility - Department of Biosciences, University of Milan) for technical assistance and support with flow cytometric analysis.

FIGURE 1

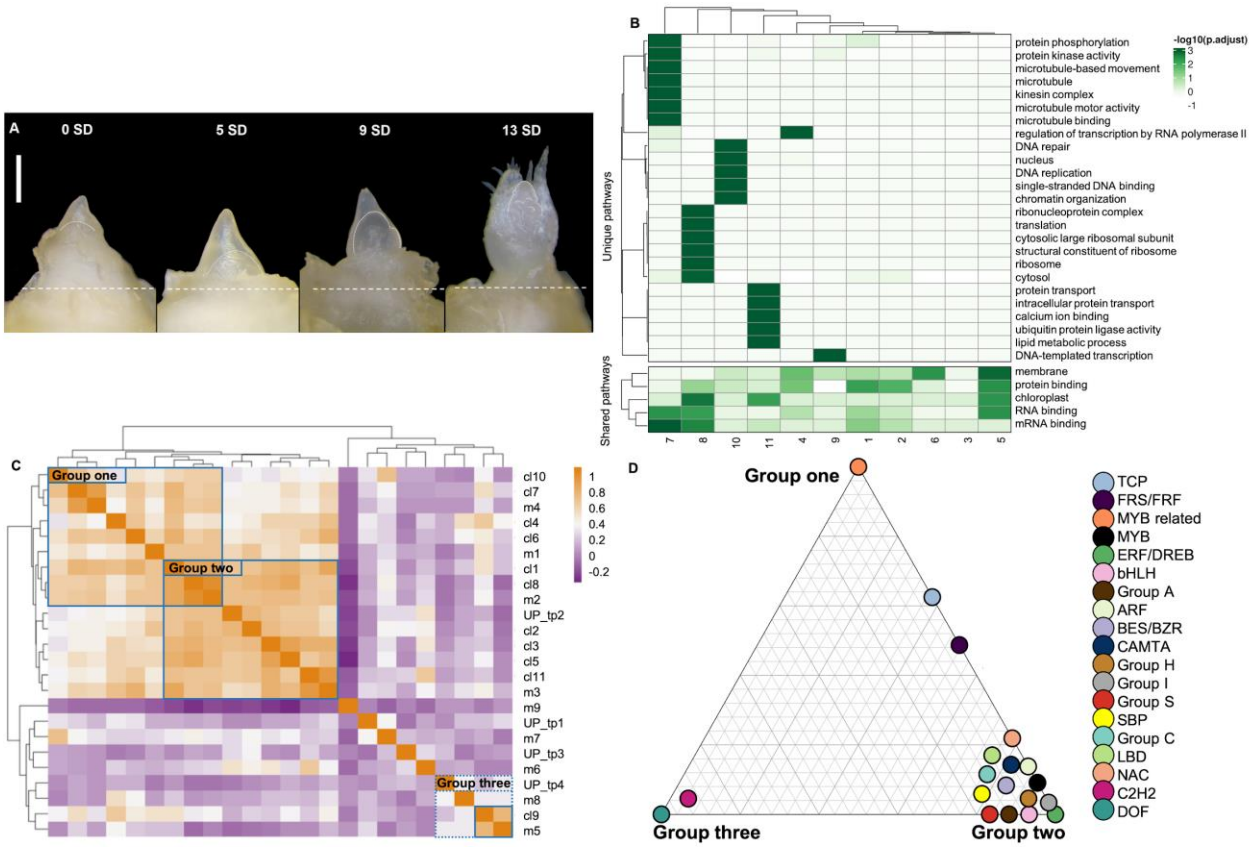

FIGURE 1 LEGEND

A) Morphological changes of the shoot apical meristems (SAM) of 6-week-old rice plants grown for six weeks under long days (LD) and then exposed to 0 short days (SD); 5 SD, 9 SD and 13 SD. Dotted curves show the SAM position behind the last meristematic leaf. Dashed lines represent the point where SAMs were cut for sampling. Scale bar: 0.5mm.

B) Heatmap of biological process gene ontology (GO) terms enriched in at least one cluster. Biological processes are indicated on the left side of the heatmap while clusters are reported on the bottom; “shared pathways” indicate GO terms enriched across all the clusters; “unique

pathways" are those that are specific to one cluster. The significance of the enrichment is expressed as –log(p.adjusted).

C) Heatmap of pairwise correlations of TFBS enrichment profiles as identified by pscan. snRNA-seq clusters are indicated by "cl", snRNA-seq pseudo time modules DEGs by "m", and bulk RNA-seq DEGs up regulated at different time points by "UP_tp". In the heatmap, colors reflect the correlation coefficient, according to the scale reported on the right. Manually inferred groups/clusters are highlighted with blue squares.

D) Ternary plot of TF families. Family and class notation according to the Jaspar database. Groups as defined in panel C, with each group arbitrarily assigned to a corner.