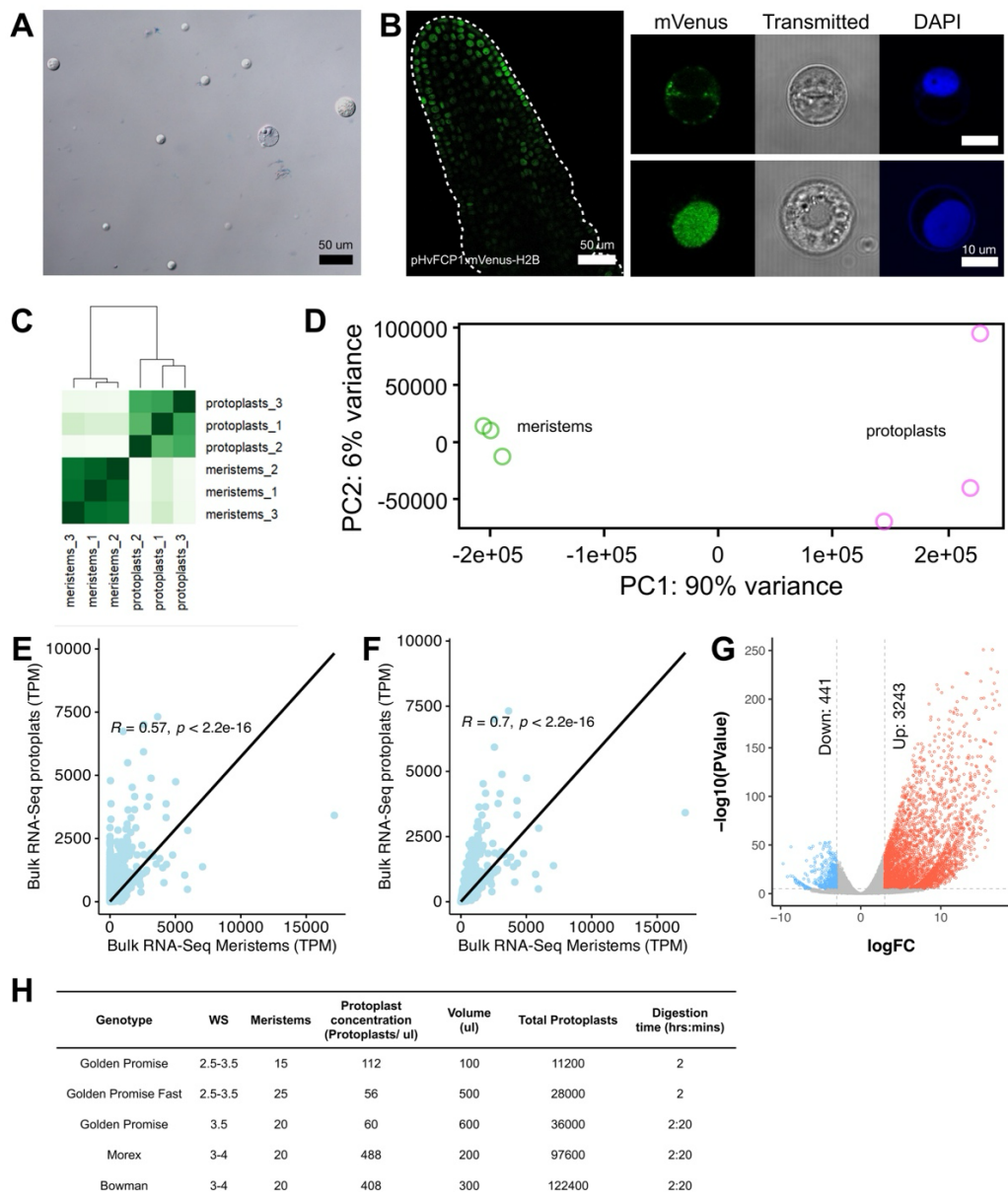


Imputation integrates single-cell and spatial gene expression data to resolve transcriptional networks in barley shoot meristem development

In the format provided by the
authors and unedited

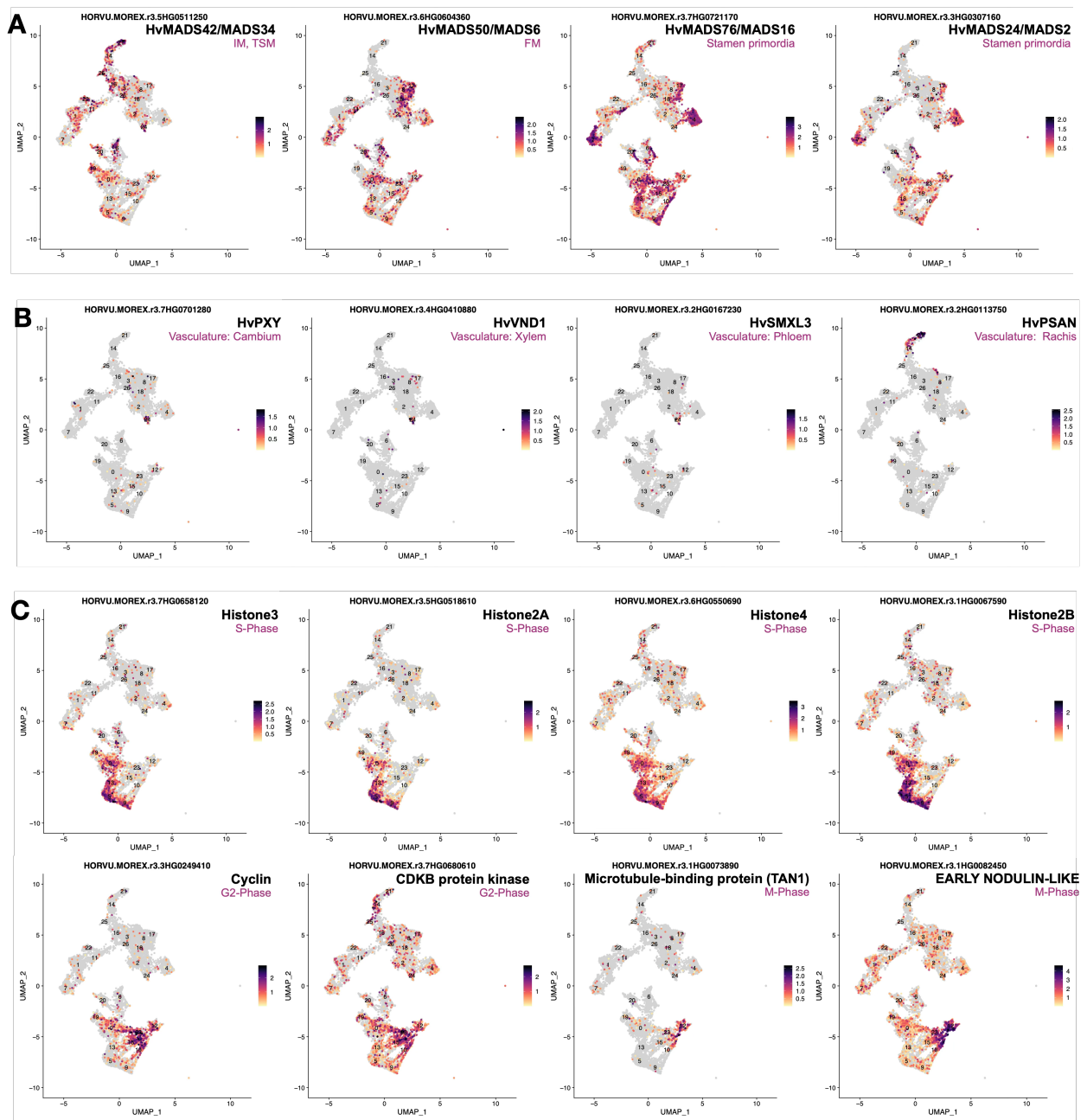
1 **Supplementary Figure1:**



2
3 **SuppFig1: Protoplast isolation and identification of differentially expressed genes during this**
4 **process.**

5 (A) Size diversity of protoplasts isolated using the protocol optimised for barley spikes. (B) The
6 transcriptional reporter line *pHvFCP1:mVenus-H2B* expressed in IM and the protoplasts recovered from
7 this line showed the presence and absence of mVenus suggesting the isolation of diverse populations of
8 protoplasts from the developing spike of barley. (C) Cluster correlation between replicates in bulk RNA-
9 Seq from intact tissue and protoplasts at W3.5. (D) Principal component analysis from bulk RNA-Seq to
10 detect differentially expressed genes during protoplasting. (E) Correlation of gene expression between bulk
11 RNA-Seq from intact spikes (meristems) and protoplasts. (F) Correlation of gene expression after removing
12 differentially expressed genes. In (E) and (F) Pearson's correlation coefficient: R. (G) Differentially
13 expressed genes with $\log_2FC > 3$ or $\log_2FC < -3$ and $FDR < 1 \times 10^{-5}$ (Supplementary Table 2). (H) Protoplast
14 recovery with different genetic backgrounds. (A) and (B) are representative images of 2 independent
15 experiments. Scale bars: 50 μm in (A) and right panel in (B), 10 μm in left panel in (B).

16 **Supplementary Figure2:**



17

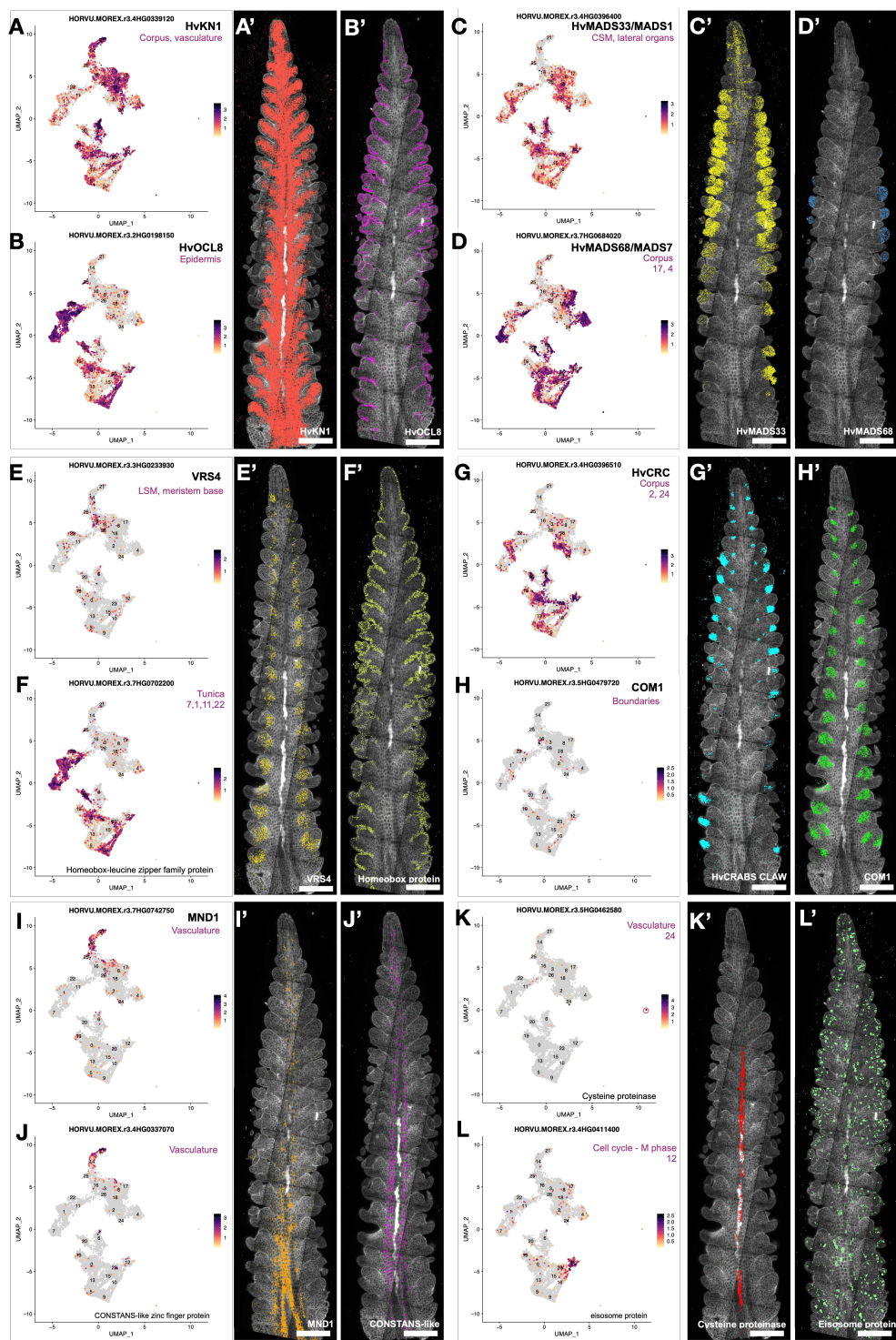
18 **SuppFig2: Marker genes specifically expressed in clusters.**

19 UMAP plots for selected marker genes associated with (A) the corpus and tunica in meristems, (B)

20 vasculature, and (C) dividing cells (two rows with graphs).

21

22 **Supplementary Figure3:**



23

24 **SuppFig3: Cluster annotation and validation using known and newly discovered markers.**

25 (A-D) Examples of mRNA expression in scRNA-Seq UMAPs and experimental localization in developing

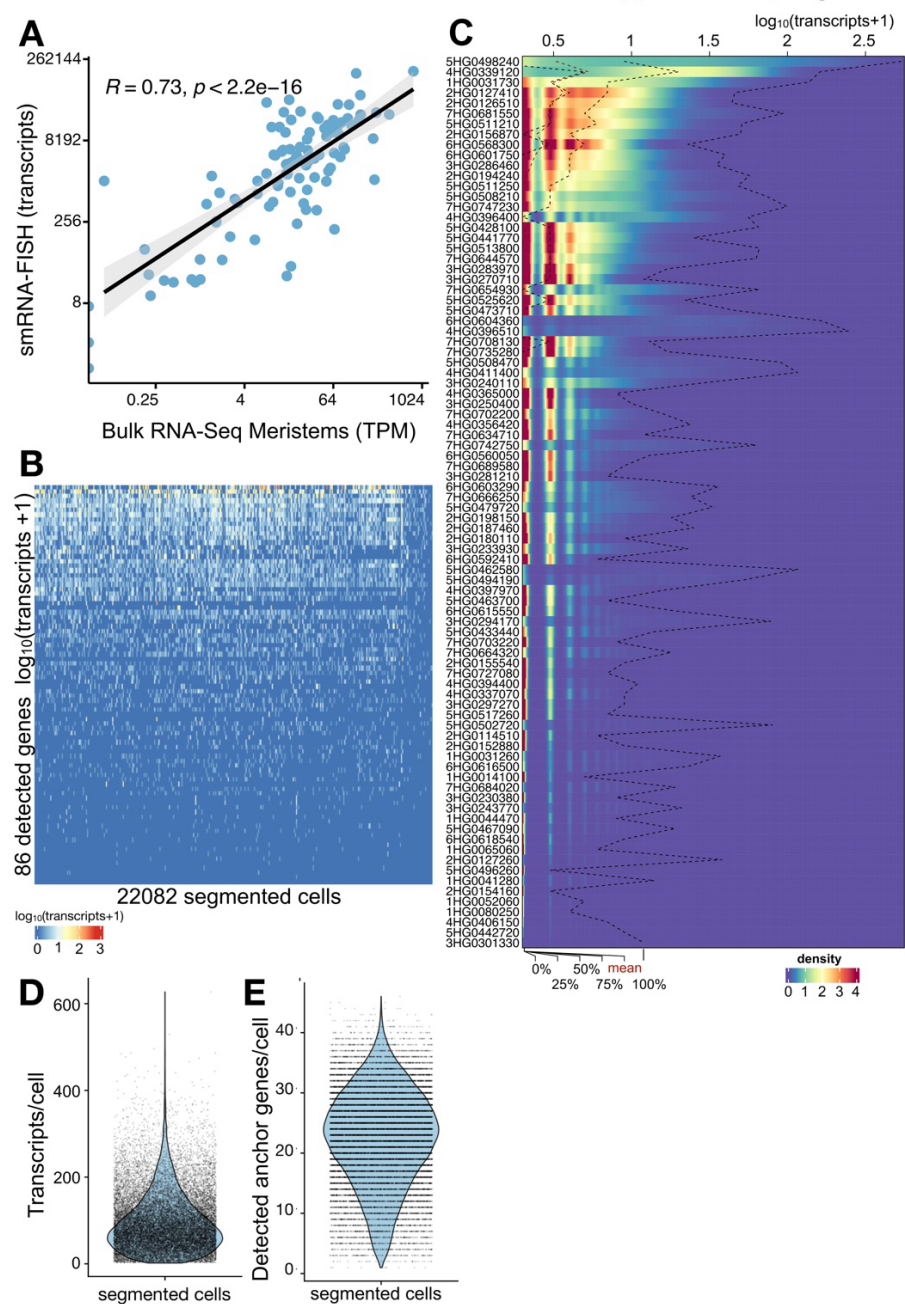
26 spikes of known markers by smRNA-FISH. (E-L) Examples of mRNA expression in scRNA-Seq UMAPs

27 and experimental localization in developing spikes by smRNA-FISH (markers newly identified by cluster

28 analysis). smRNA-FISH sections are representative from at least 5 independent developing spikes. Scale

29 bars: 200 μ m.

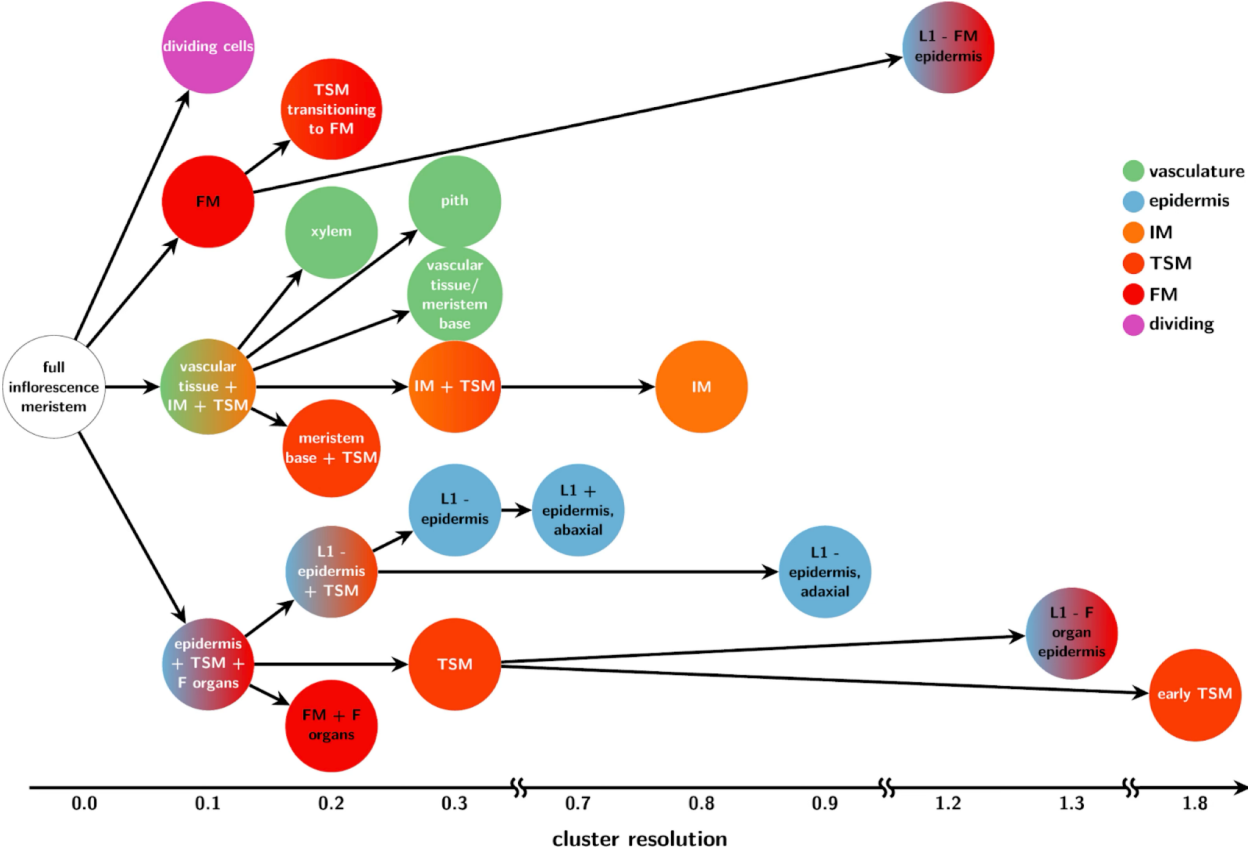
30 **Supplementary Figure4:**



SuppFig4: smRNA-FISH is quantitative and correlates with detected gene expression.

(A) smRNA-FISH transcripts (counts) are linearly correlated with the genes detected by bulk RNA-Seq. (B) The 86 informative genes (SuppTable4) detected by smRNA-FISH show a high dynamic range of transcript expression (counts) across the segmented cells from tissue at W3.5. (C) Frequency density distribution of counts detected in the segmented cells from tissue at W3.5. (D) Range of total transcripts detected in the 22,082 segmented cells. (E) Range of genes (from the 86 used for smRNA-FISH) detected in the segmented cells. In (A), the x and y axes were log₂-transformed to show the correlation more clearly. In (B) and (C), the transcripts are expressed as $\log_{10}(\text{transcripts} + 1)$ to simplify the visualization of the high dynamic range between genes.

Supplementary Figure5:

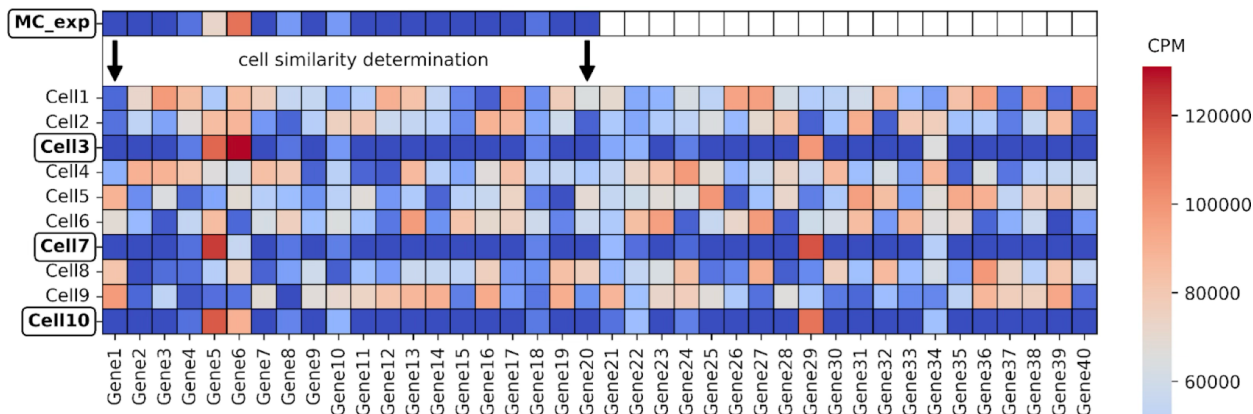


SupplFig5: Overview of the cell types emerging in the inflorescence meristem depending on the cluster resolution.

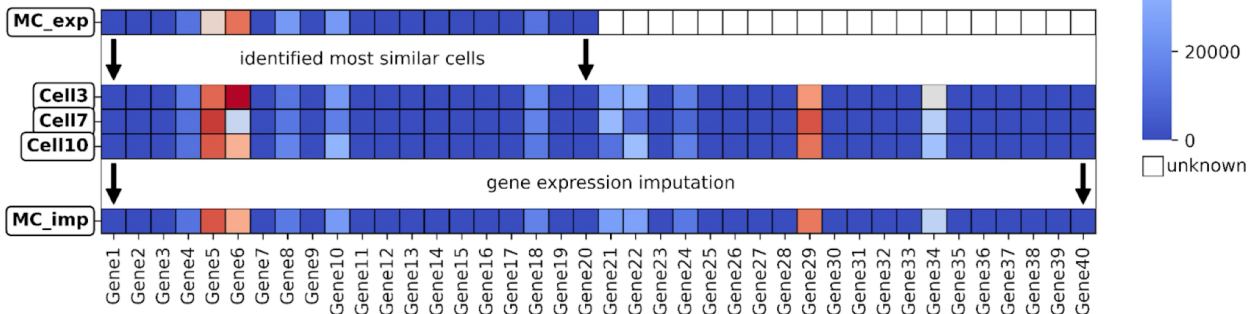
Cell types become distinct depending on clustering resolution parameters. Clusters that separate out at low resolution have very distinct gene expression patterns, whereas those appearing only at high resolutions of the *Seurat* clustering algorithm show less distinct gene expression profiles (IM, inflorescence meristem; TSM, triple spikelet meristem; FM, floret meristem). The color code shows the main cell lineages identified in the plot under specific cluster resolutions.

Supplementary Figure6:

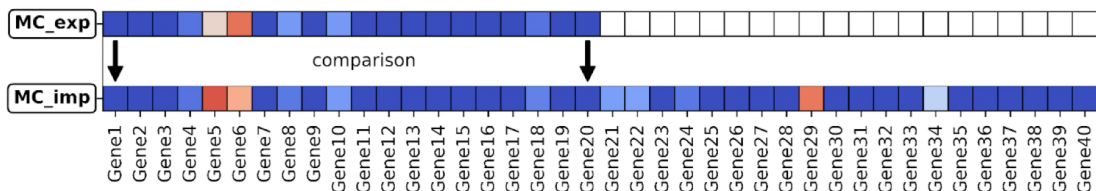
A



B



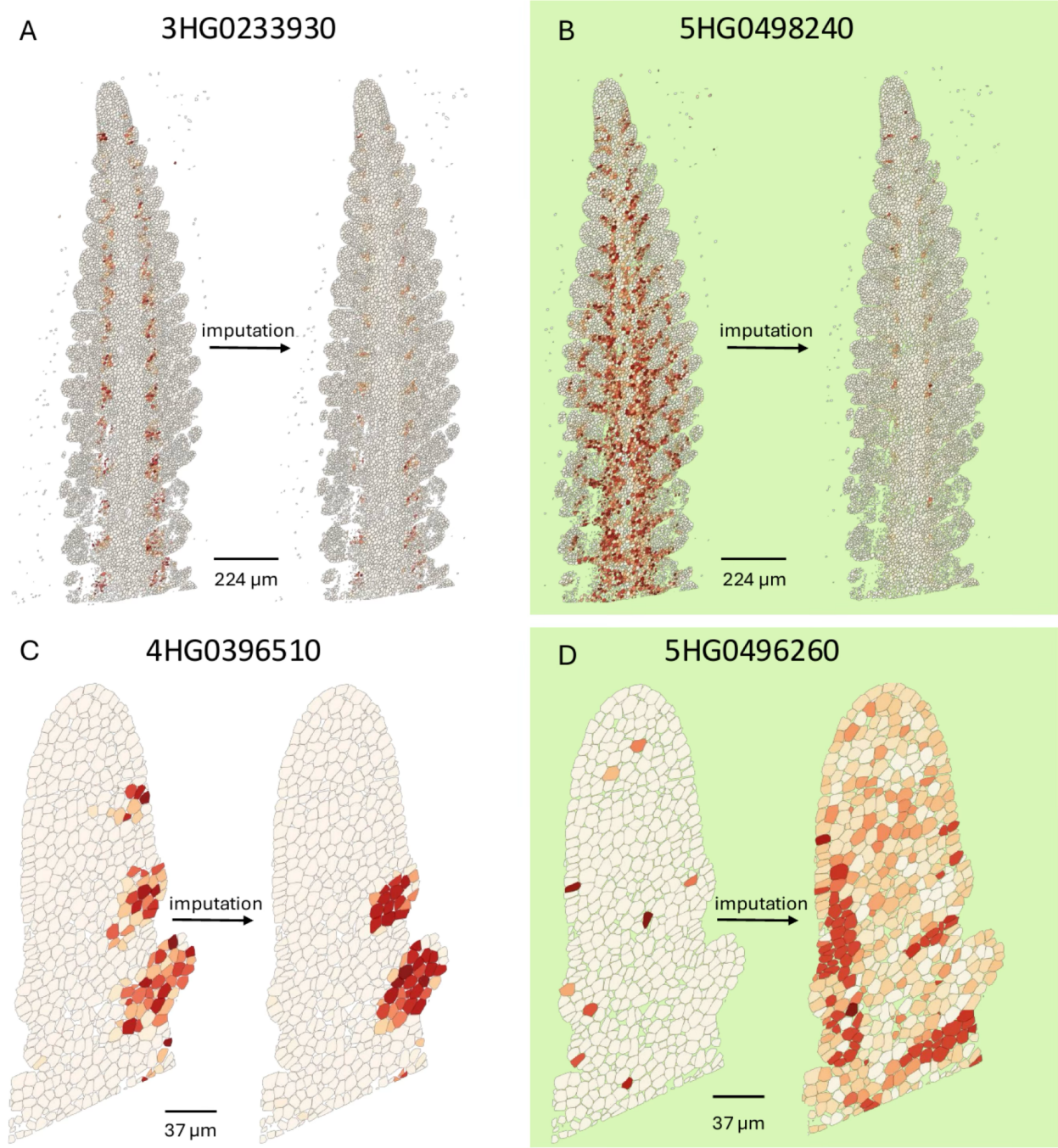
C



SupplFig6: Scheme of the gene expression imputation workflow.

(A) The first step of gene expression imputation is the determination of the most similar scRNA-Seq cells for each cell that was experimentally measured by molecular cartography (MC_exp) based on their gene expression, using the genes from the MC (Gene1-Gene20). Cell similarity is determined by cosine similarity (CS). The most similar cells are highlighted on the left. (B) Gene expression is determined by calculating the weighted average of the gene expression from the most similar cells, using each cell's CS as weight. This yields an imputed version of the same cell from the molecular cartography (MC_imp), containing gene expression information about all genes measured in the scRNA-Seq experiment. (C) Validation by comparing the experimentally determined and computationally imputed version of each MC-cell, using the available genes measured in the MC (Gene1-Gene20). Similarity between imputed and experimentally determined gene expression values was calculated using CS.

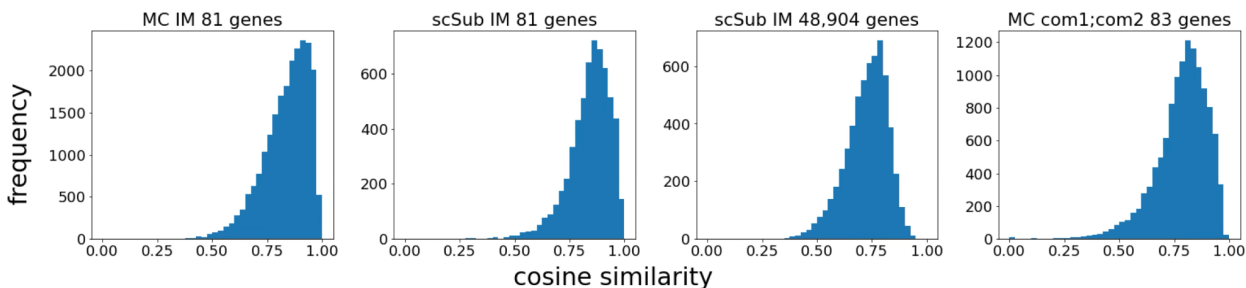
66 **Supplementary Figure7:**



67
68 **SupplFig7: Comparison of measured and imputed gene expression**

69 (A, B) The measured and imputed gene expression patterns and levels for two genes in the inflorescence
70 meristem. (C, D) The measured and imputed gene expression patterns and levels for two genes in the
71 vegetative shoot apical meristem. The green background indicates that the genes displayed in (B, D) are
72 affected by the protoplasting procedure, resulting in different patterns.
73

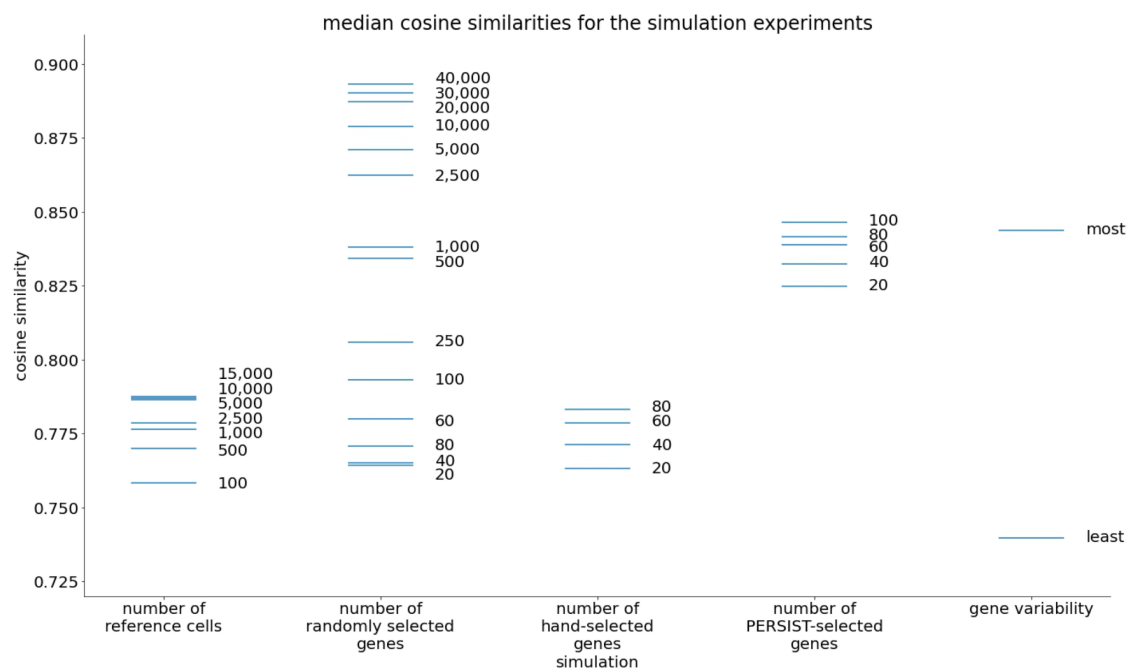
Supplementary Figure8:



SupplFig8: Gene expression imputation quality assessment.

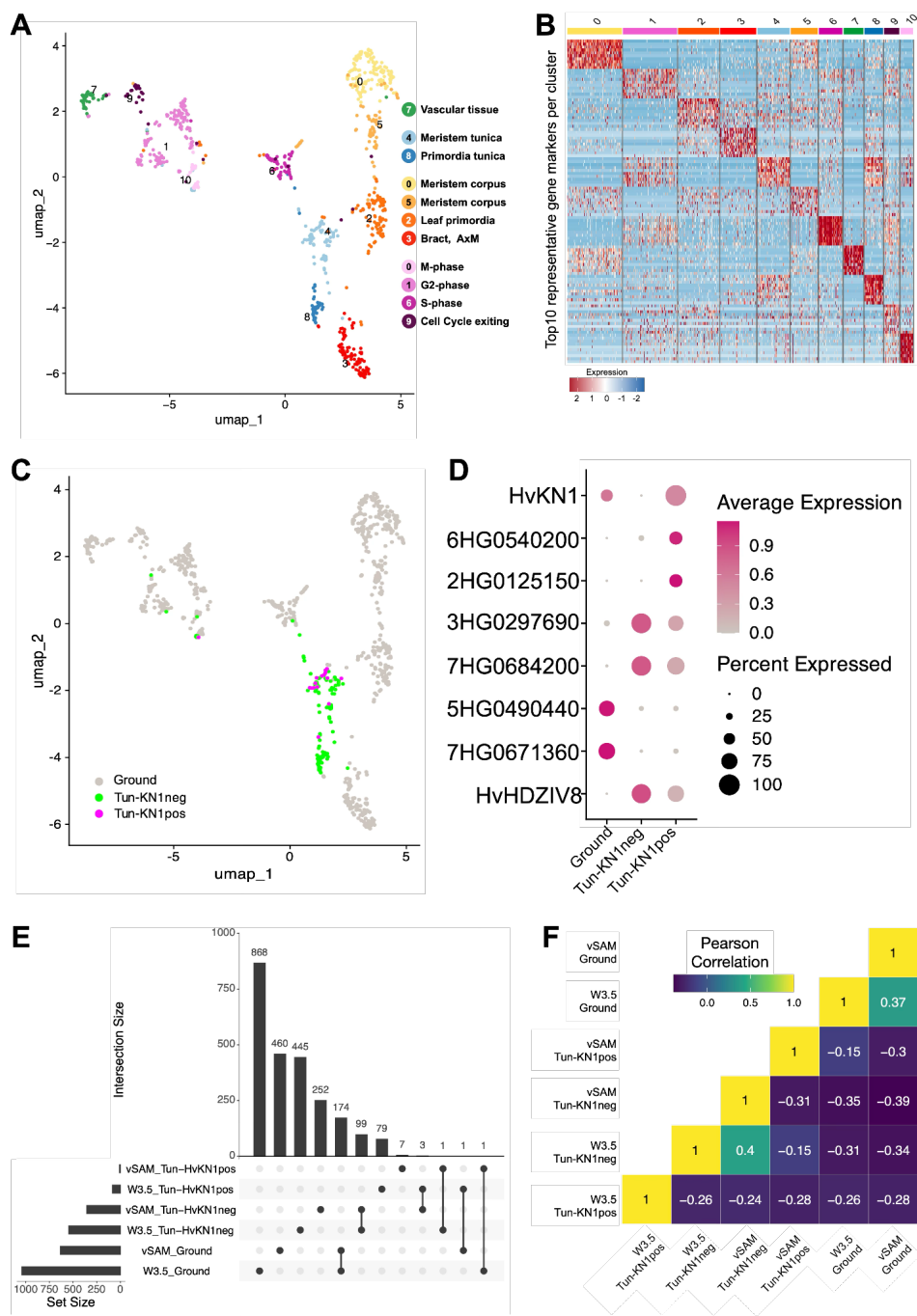
Imputed and experimentally determined gene expression values were compared for four different datasets using cosine similarity (CS) for each cell in the test dataset. The first histogram shows the CS for all 22,082 cells in the inflorescence meristem (IM) Molecular Cartography (MC) dataset for the 81 experimentally determined genes. The second histogram shows the CS for all 5,910 cells in the subset of the scRNA-Seq dataset for the same 81 genes available in the MC. The third histogram shows the CS for the same 5,910 cells in the subset of the scRNA-Seq dataset for all 48,904 genes experimentally determined in the scRNA-Seq experiment. The fourth histogram shows the CS for all 12,092 cells in the *com1a;com2g* MC dataset for the 83 experimentally determined genes.

86 **Supplementary Figure9:**



87
88
89
90
91
92
93
94
95

SupplFig9: Gene expression imputation quality simulation.
Imputation quality for five different simulations with different numbers of reference cells, for different numbers of genes selected for the similarity determination, either by random selection, hand selection of known marker genes and genes of interest, by selection with the *PERSIST* tool, or by selecting the top and bottom 250 most and least variable genes. Each line represents the median cosine similarity score from the validation, which compares imputed gene expression patterns to the experimentally measured patterns. The displayed medians each represent a mean value from n = 5 simulation experiments.

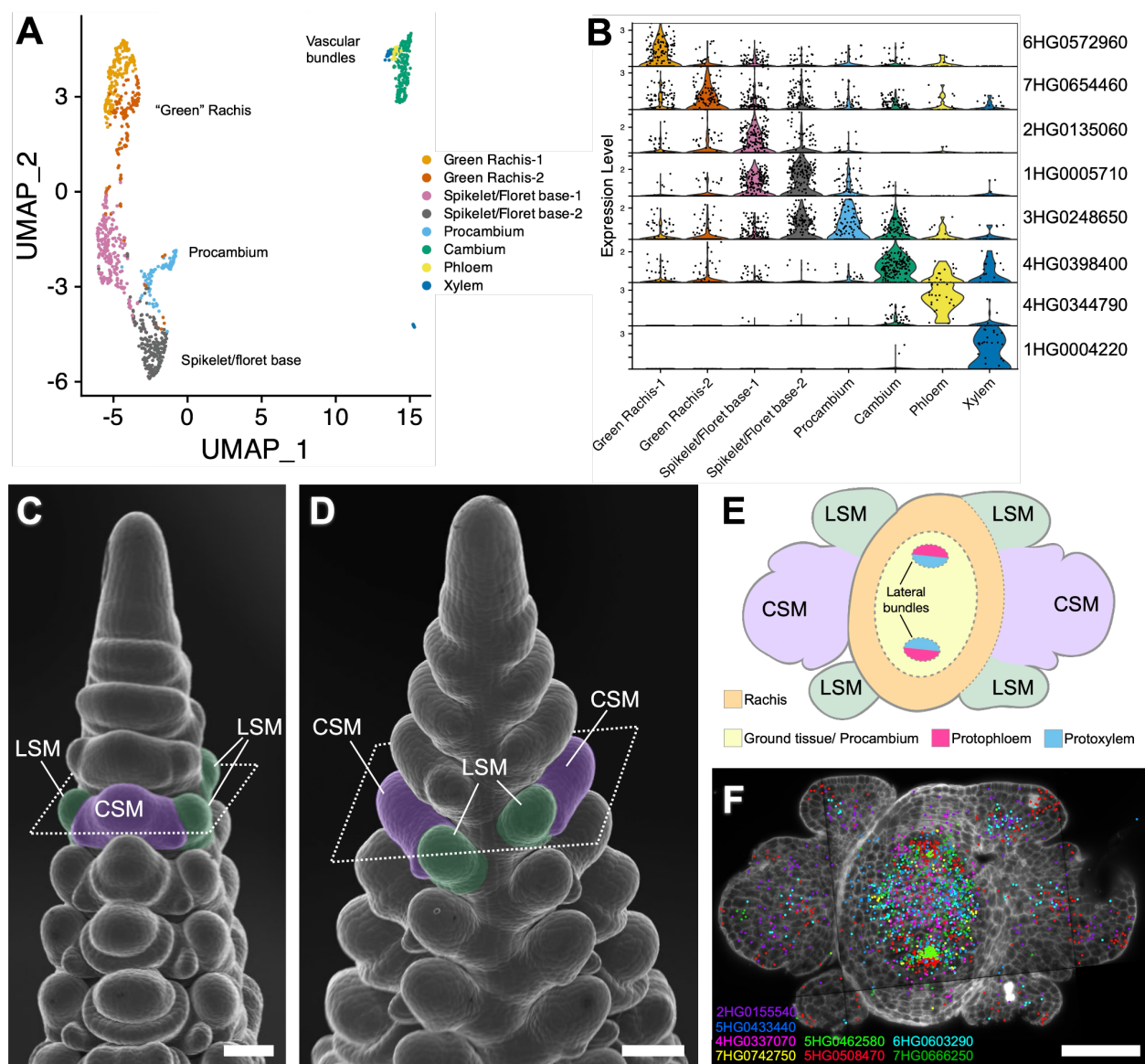


97

98 **SupplFig10: Single-cell atlas of barley vegetative SAM.**

99 (A) UMAP dimensional reduction of scRNA-Seq in the vegetative SAM (vSAM). (B) Top 10 marker genes
100 per cluster prioritized by P_{adj} . (C) *HvKN1* expression in tunica cells. (D) Selected marker genes in tunica
101 cells expressing *HvKN1* in the vSAM. (E) UpSet plot showing the number of differentially expressed genes
102 (DEGs) specifically identified or shared between cell populations: tunica expressing *HvKN1*, tunica lacking
103 *HvKN1* expression, and ground tissue (corpus), each for vSAM and W3.5. Set size indicates the total
104 number of DEGs identified in each group of cells. Intersection size shows the number of DEGs shared
105 between the indicated cell groups. (F) Pearson correlation between the number of DEGs identified and the
106 number of shared DEGs comparing all different cell populations. Scale bars: 50 μ m.

107 **Supplementary Figure11:**



SupplFig11: Subcluster analysis in vasculature.

(A) UMAP plot with subcluster organization in vascular cells. (B) Violin plots of selected markers expressed in subclusters. (C) Scanning electron micrograph frontal view of a spike, indicating the position of the transverse section (dashed lines) used in multiplex smRNA-FISH (E,F) for vascular cells (LSM in green, CSM in violet). (D) Lateral view of the same spike. (E) Schematic representation of transversal section at the position indicated in (C) and (D). (F) Selected expression profiles in vascular tissues based on multiplex smRNA-FISH. smRNA-FISH sections are representative from at least 5 independent developing spikes. Scale bars: 100 μ m.

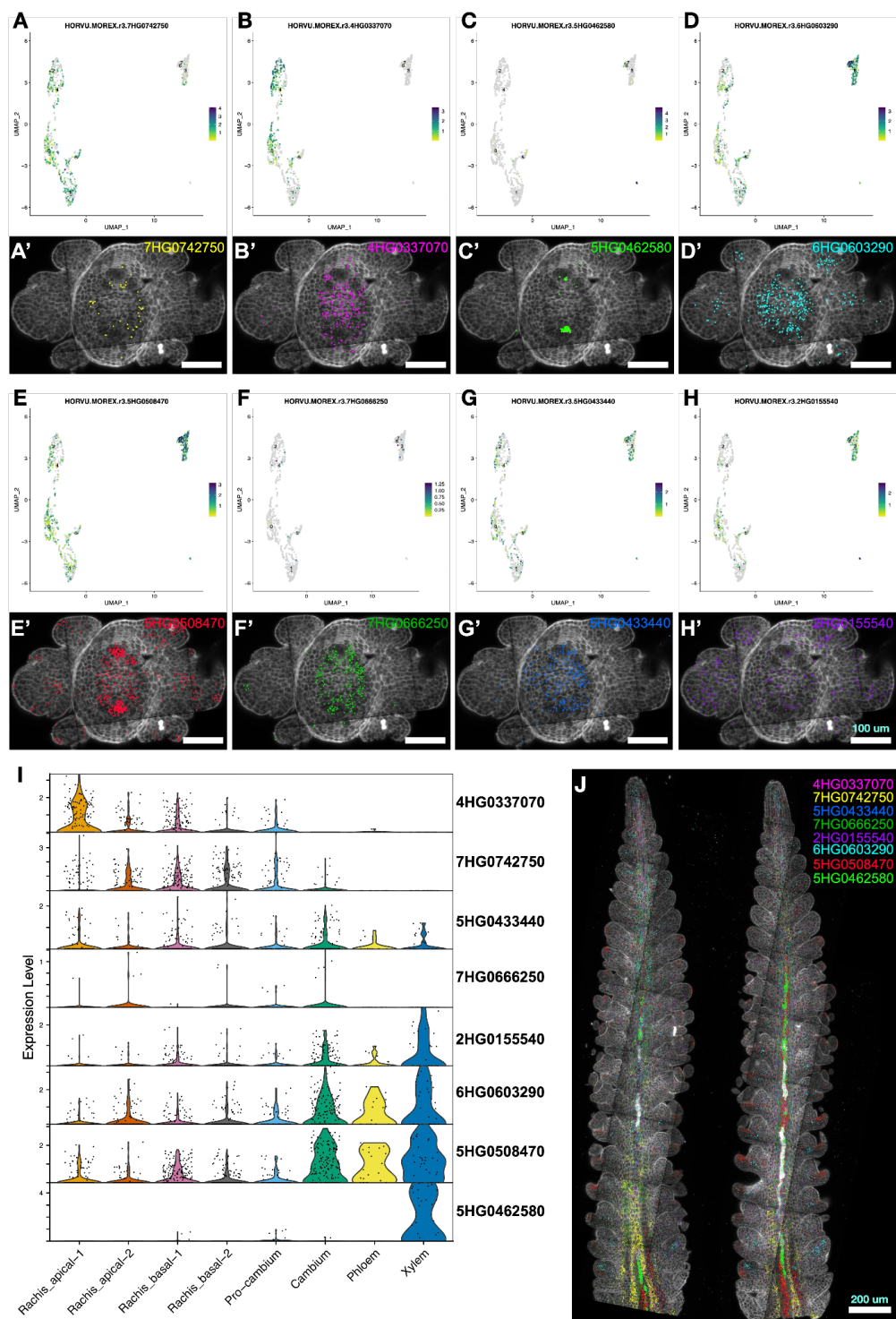
Supplementary Data:

Identification of cluster and marker genes for spike greening

The total number of grains produced on a spike depends on the number of SMs that are initiated, and on their maintenance. SM number is largely controlled by flowering time genes that set heading time, whereas the development of the vasculature, light signals and activation of chloroplasts promotes SM survival⁹¹. In developing spikes of the subfamily Pooideae, development of plastids and the photosynthetic machinery in the rachis can be observed as an acropetal gradient of green cells in the vasculature. Using the identified clusters related to vascular tissues, we analysed these cells by unbiased subclustering (**SupplFig11**, **SupplFig12**). We identified marker genes such as *Thioredoxin z* (*HvTRX*, 4HG0341820) involved in plastid division, and *PHOTOSYSTEM II LIGHT HARVESTING COMPLEX GENE 2* (*HvLHCB2*;3, 5HG0499400) and Photosystem I reaction center subunit N (*HvPSAN*, 2HG0113750), both involved in photosynthesis and important for green rachis pigmentation.

At W3.5, the developing spike forms lateral bundles in the rachis, which will connect the vascular tissues to the spikelets. In our subclusters, we can recognize the procambium in the ground tissues and the cambium differentiating into protoxylem and protophloem, patterning the lateral bundles. We identified additional protophloem and protoxylem markers. Among the new markers for developing barley phloem, we found *HvBAM5* (4HG0344790), DNA-BINDING WITH ONE FINGER (DOF) transcription factor genes (5HG0441840, 5HG0526260) and two homologs of Arabidopsis *ALTERED PHLOEM DEVELOPMENT* (6HG0570890, 7HG0738570), and for the developing xylem we found the CLE-peptide homolog of FON2 SPARE1 in barley (*HvFOS1*, 1HG0008860), six laccases (7HG0667290, 4HG0357240, 3HG0302570, 1HG0073500, 2HG0135650, 3HG0302520), two LOGs (4HG0408000, 1HG0043530), and two basic helix-loop-helix transcription factors (7HG0731430, 5HG0484930) related to *PERICYCLE FACTOR TYPE-A 2* (*PFA2*) and *TARGET OF MONOPTEROS 5*, respectively, already associated with xylem differentiation in *Pinus* and Arabidopsis (**SupplTable13**).

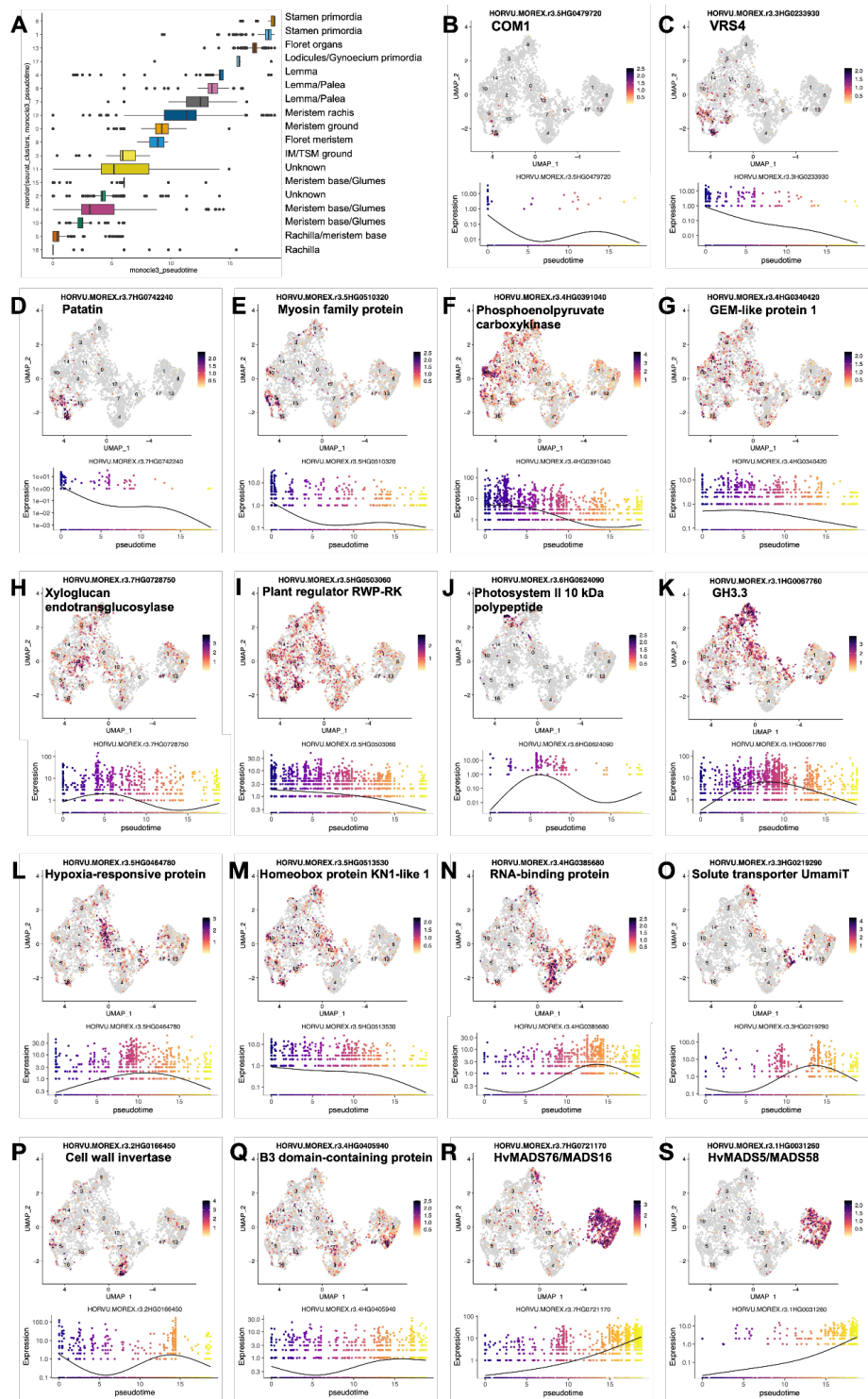
147 **Supplementary Figure12:**



148 **SupplFig12: Subcluster analysis and validation in vascular tissues**

149 (A-H) UMAPs in subcluster analysis using clusters related to vascular tissue and their smRNA-FISH
150 expression profiles in transverse sections of W3.5 developing spikes in barley. (I) Violin plot of gene
151 expression in subclusters. (J) Expression profiles in sagittal sections of developing spikes. smRNA-FISH
152 sections are representative from at least 5 independent developing spikes. Scale bars: (A')–(H') 100 μ m;
153 (J) 200 μ m.

155 **Supplementary Figure13:**

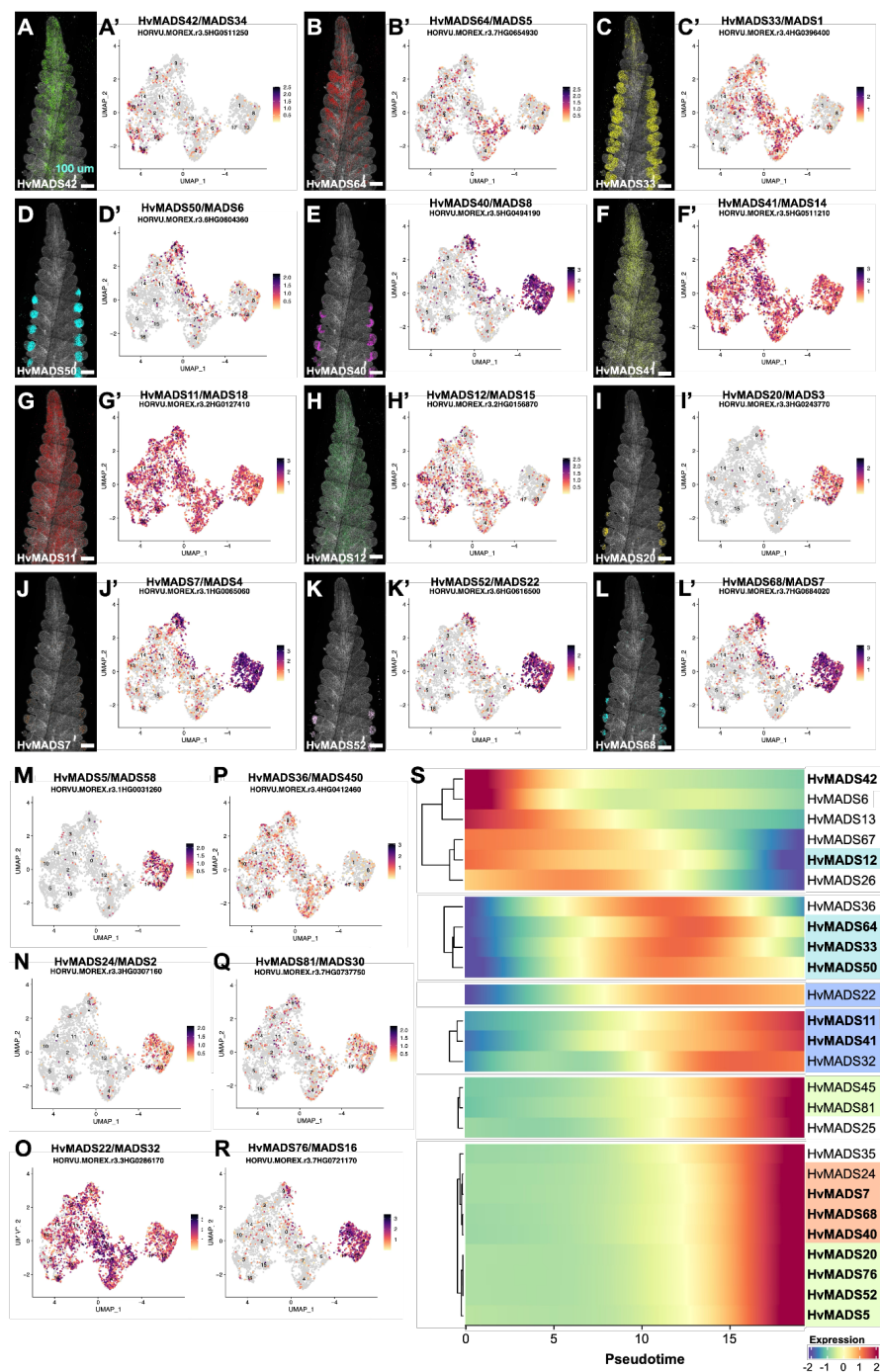


156

157 **SuppIFig13: Subcluster analysis in corpus cells and their modulation across the pseudotime.**

158 (A) Corpus subclusters arranged along pseudotime. (B) Expression levels of *COM1* in corpus subclusters
159 and its modulation in pseudotime. (C) Expression levels of *VRS4* in subclusters and its modulation in
160 pseudotime. *COM1* and *VRS4* are early markers for spikelet meristem formation. (D)-(S) Expression
161 profiles and their modulation along pseudotime for selected marker genes in corpus subclusters.

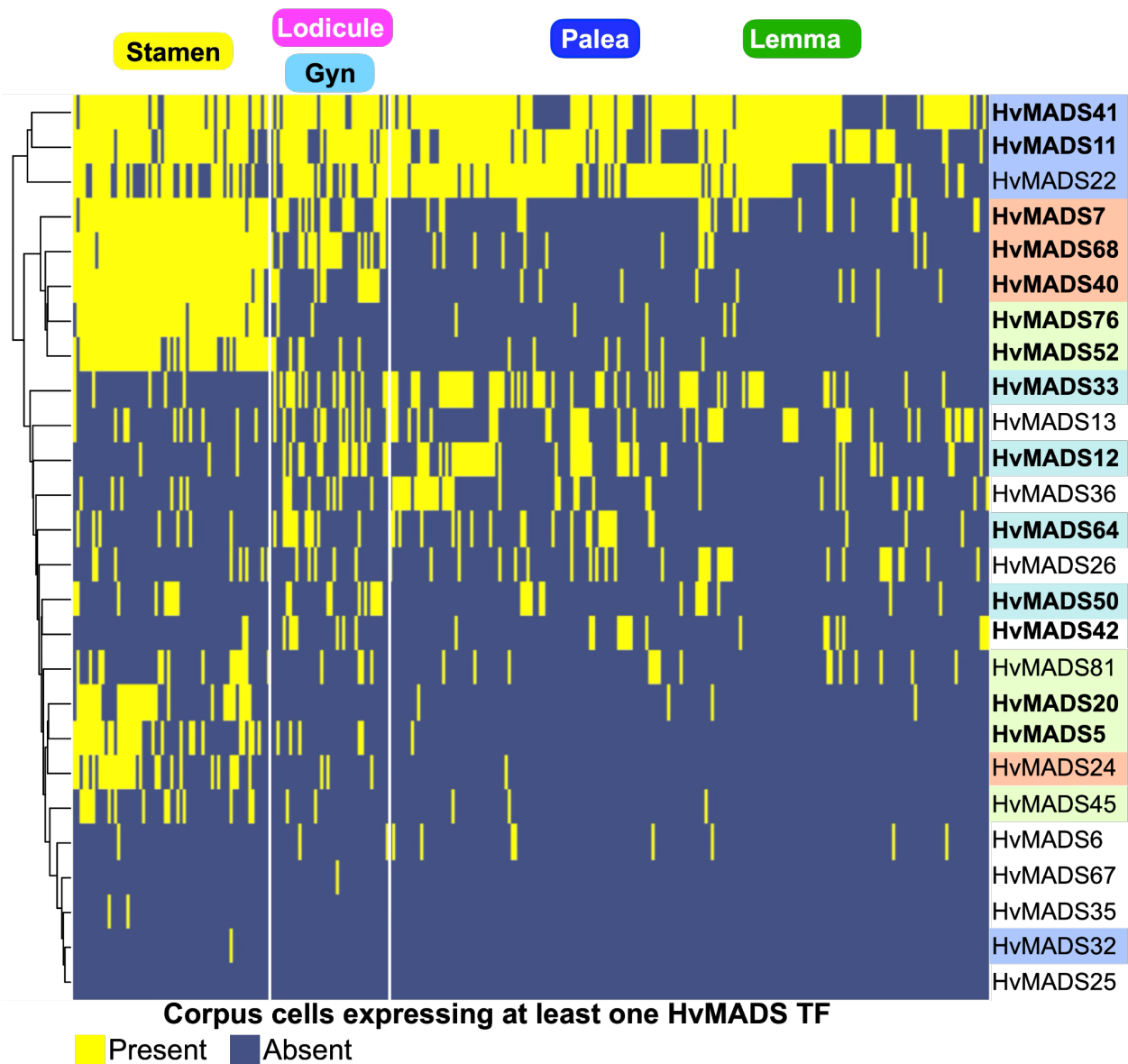
Supplementary Figure14:



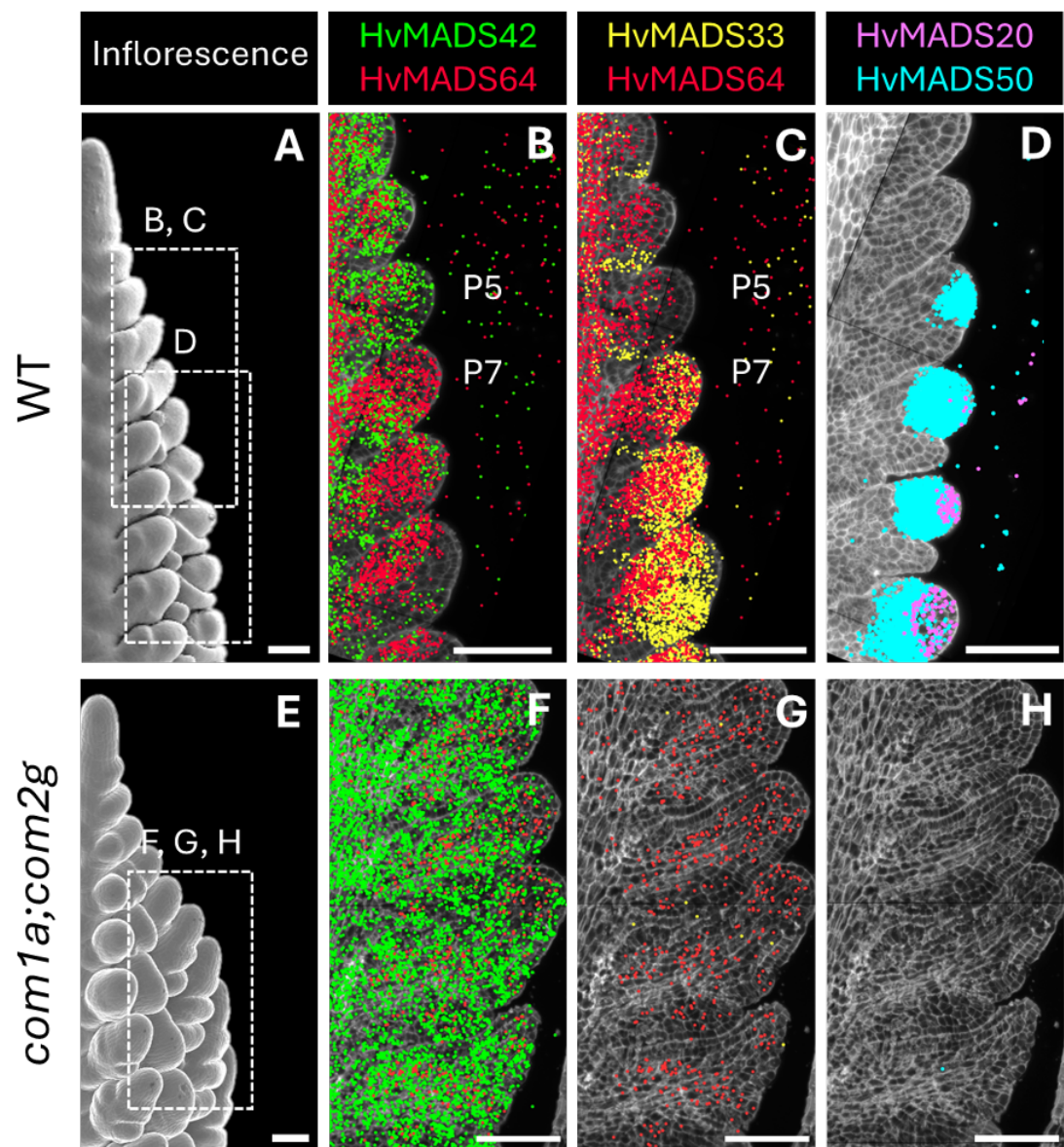
SuppIFig14: MADS-box transcription factor genes expressed in corpus subclusters and their modulation along pseudotime

(A)–(E) MADS-box smRNA-FISH expression patterns and their expression profiles in corpus subclusters. (F)–(L) Additional expression patterns of selected MADS-box genes. (M)–(R) Additional UMAP plots of MADS-box genes expressed in corpus subclusters. (S) Modulation along pseudotime of MADS-box transcription factor genes expressed in corpus subclusters. MADS-box genes with smRNA-FISH data reported in this study are shown in bold. Assignment of MADS-box classes: A, lilac; B, peach; C, pale green; E, pale blue. smRNA-FISH sections are representative from at least 5 independent developing spikes Scale bars: 100 μ m.

Supplementary Figure15:



SupplFig15: MADS-box transcription factor genes co-expressed in corpus cells.
The presence or absence of MADS-box transcription factor genes detected by scRNA-Seq in corpus cells, hierarchically ordered by the MADS-box gene expressed in each cell. Assignment of MADS-box classes: A, lilac; B, peach; C, pale green; E, pale blue. The origin of the putative floret organ of the cells is indicated above the heat map, according to the HvMADS class.



SupplFig16: Stage-specific MADS-box transcription factor genes expressed in WT and *com1a;com2g* inflorescences.

(A) Scanning electron micrograph of WT barley inflorescence at W3.5. Dashed areas indicate the location of primordia displayed in panels B–D. (B,C) Longitudinal sections of developing WT spikelets: smRNA-FISH data showing the localisation of *HvMADS42* (green), *HvMADS64* (red), and *HvMADS33* (yellow) transcripts. (D) Developing WT florets: smRNA-FISH data showing the localization of *HvMADS20* (purple) and *HvMADS50* (cyan) transcripts. (E) Scanning electron micrograph of *com1a;com2g* branched inflorescence. The segmented square indicates the location of IM-like branches displayed in panels F–H. (F–H) Longitudinal sections of *com1a;com2g* IM-like branches: smRNA-FISH data showing the localization of *HvMADS42* (green), *HvMADS64* (red), *HvMADS33* (yellow), *HvMADS20* (purple), and *HvMADS50* (cyan) transcripts. smRNA-FISH sections are representative from at least 2 independent developing spikes. Scale bars: 100 μ m.