Single-Cell RNA-seq based elucidation of the antidepressant hyperforin biosynthesis *de novo* in St. John's wort.

Song Wu[1,2], Ana Luisa Malaco Morotti[1], Jun Yang[1], Ertao Wang[1] and Evangelos C. Tatsis[1,3*]

[1]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences; Shanghai, 200032, China

[2]University of Chinese Academy of Sciences; Shanghai, China

[3]CEPAMS - CAS–JIC Centre of Excellence for Plant and Microbial Science, Shanghai, China

*Corresponding author. Email: etatsis@cemps.ac.cn

Summary

Hyperforin is the compound responsible for the effectiveness of St. John's wort (*Hypericum perforatum*) as an antidepressant, but its biosynthesis remains unknown. Gene discovery based on co-expression analysis of bulk RNA-sequencing data or genome mining failed to discover the missing steps in hyperforin biosynthesis. Here we sequenced the tetraploid *H. perforatum* genome. By single-cell RNA-seq, we identified a distinct type of cells, Hyper cells, wherein hyperforin biosynthesis *de novo* takes place. Through pathway reconstitution in yeast and tobacco, we identify and characterize four transmembrane prenyltransferases to resolve hyperforin biosynthesis. The hyperforin polycyclic scaffold is created by a reaction cascade involving an irregular isoprenoid coupling and a tandem cyclization. Our findings reveal how and where hyperforin is biosynthesized that enables synthetic-biology reconstitution of the complete pathway. These results deepen our comprehension of specialized metabolism at the cellular level, and we anticipate acceleration of pathway elucidation in plant metabolism.

Main text

St. John's wort (SJW, *Hypericum perforatum*, Hypericaceae) is a high valued medicinal plant with annual global sales exceeding US$6 billion[1]. In classical antiquity, extracts from aerial SJW organs were used to treat skin ailments due to antiseptic and anti-inflammatory properties[2], while nowadays are prescribed to treat mild to moderate depression, associated with fewer adverse reactions over common antidepressants[3,4]. Its primary potent bioactive, hyperforin, acts similarly to traditional antidepressants by inhibiting neurotransmitter re-uptake like serotonin, though via a different mechanism[5-9]. Unlike conventional antidepressants, hyperforin induces Ca^{2+} and Na^{+} by uniquely activating TRPC6 receptors, disrupting cation gradients and reducing neurotransmitter uptake. This unique process makes hyperforin a promising agent for treating depression[10].

H. perforatum is endemic in Eurasia and North Africa and introduced in the Americas and Oceania by human migration. It originates from an ancient hybridization event in Siberia between *H. maculatum* and *H. attenuatum*[11]. Typically tetraploid, with a chromosome number of 2n = 4x = 32[11], the plant also reproduces asexually through facultative apomixis, leading to populations with diploid, triploid and hexaploid cytotypes[12]. Besides hyperforin, SJW also biosynthesizes diverse specialized metabolites including essential oils rich in complex terpenoids[13], flavonoids, biflavonoids, phenolic acids and naphtodianthrones[14].

Hyperforin is the first polycyclic polyprenylated acylphloroglucinol (PPAP) to be isolated[15] (Figure 1A), a diverse class of natural products with various biological activities that are taxonomic markers for the family Hypericaceae[16-18]. Hyperforin's unique structure features an isoprenyl group at C8 (Figure 1A), thus poses a challenging task for enantioselective synthesis, inspiring chemists to pursue novel organic-synthesis strategies[16-18]. Such synthesis routes involve multiple steps can result in total yield of 1.64%[19]. Despite hyperforin's significance, its biosynthesis *de novo* is unresolved. With reconstitution of the pathway in microbial or plant synthetic-biology systems,

obtaining hyperforin from such sources is a sustainable and environmentally superior strategy to isolate this valuable pharmacological agent.

To date, details of the biosynthesis of specialized metabolites in SJW, and for the majority of the hyperforin biosynthesis steps, are sparse. We have characterized three enzymes, namely a branched-chain amino-acid dehydrogenase (BCKDHA), a CoA ligase (CCL) and an aromatic polyketide synthase (PKS) from two biosynthetic gene clusters (*HpBGC1–2*) that encode the first three steps dedicated to biosynthesis of the polyketide hyperforin precursor — phloroisobutyrophenone (PIBP)[20] (Figure 1A). Provision of ^{13}C-labeled substrates to *H. perforatum* seedlings revealed that the rest of the hyperforin carbon atoms comprise five isoprenoid units derived from the plastid-localized methyl-*D*-erythritol-4-phosphate (MEP) pathway[21] (Figure 1A). Chemical logic suggests that the isoprenoid units derived from the plastid localized MEP pathway[21-23] are adjoined to the polyketide PIBP by electrophilic substitution, involving an additional ring closure to biosynthesize the bicyclo[3.3.1]nonane scaffold (Figure S1) of hyperforin. Transfer of isoprenoid moieties from pyrophosphate donor substrates to aromatic acceptor substrates in plants is typically catalyzed by membrane-bound aromatic prenyltransferases [24-27].

High-quality genome assemblies are of paramount importance for complete pathway elucidation of metabolites with complicated structures[28,29]. Two fragmented and incomplete annotations of a diploid *H. perforatum* genome are available [30,31], but these are insufficient for hyperforin pathway elucidation (Figure S2).

Here, to fully elucidate the hyperforin biosynthetic pathway, we sequenced the tetraploid genome of *H. perforatum* and generated single-cell atlases for leaves and flowers. Based on mapping of previously characterized upstream genes, we identified a new type of cells — Hyper cells — wherein biosynthesis *de novo* of hyperforin occurs. By using two synthetic-biology platforms for heterologous expression (yeast and tobacco), we discover and characterize four new UbiA-family prenyltransferases that complete the hyperforin biosynthesis pathway. The last step for hyperforin biosynthesis involves an unusual prenyltransferase catalyzing an irregular branching prenylation and a rare tandem cyclization.

Results

Sequencing and assembly of the tetraploid *Hypericum perforatum* genome

To enable the discovery of the missing steps in hyperforin biosynthesis, we generated transcriptome and genome-sequence resources *de novo* for *H. perforatum* using a suite of Illumina, PacBio HiFi and Hi-C sequencing (Table S1). We used Illumina paired-end short-read sequencing to survey the genome, and based on k-mer analysis, the genome size was estimated as 1.71 Gb (Table S2). PacBio HiFi long-read sequencing was mainly used for genome assembly in conjunction with Illumina sequencing for contig error correction, resulting in a 1.54-Gb (Table S3) draft assembly comprising 3,710 scaffolds and an N50 value 42.44 Mb. A Hi-C (*in vivo* fixation of chromosomes) library refined the first version of the reference genome and the scaffolds aligned into 32 groups

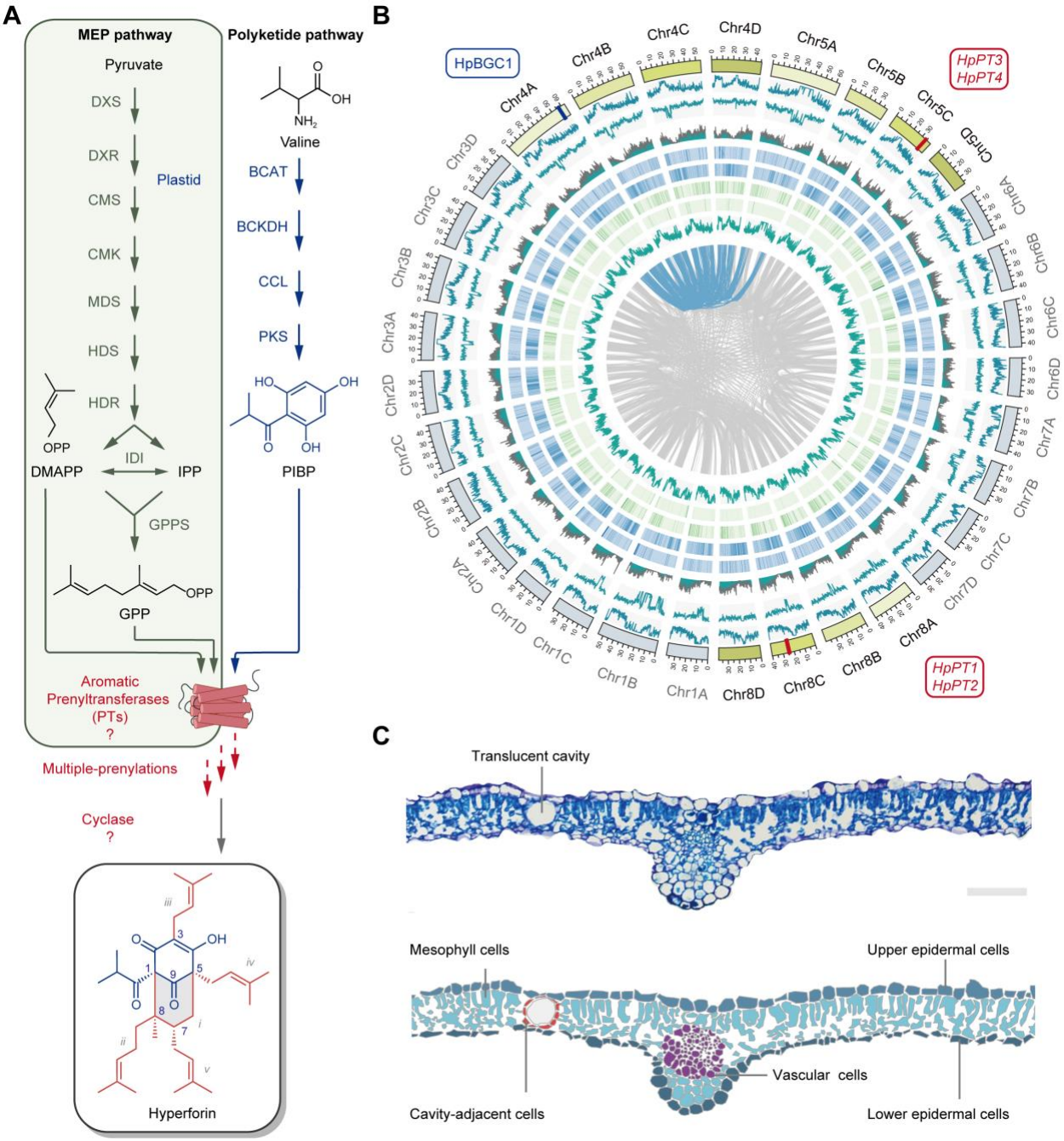
86 hereafter referred to as chromosomes (Figure 1B, Table S4, Figure S3), which represented a
87 tetraploid accession with a haploid chromosome number of 8. Coverage of the final assembly was
88 evaluated by Illumina short-read mapping, CEGMA and BUSCO (Tables S5–7) that indicate a high-
89 quality assembly with 97.6% completeness. A total of 98,005 genes (Table S8) were annotated by a
90 pipeline combining *de novo* and homology-based predictions, with bulk RNA-seq data[32].
91 Transposable elements account for 56.95% of the genome (Table S9).

92 Functional annotation of genes was performed by searching against publicly available databases
93 including SwissProt, NCBI-NR, Pfam, KEGG and InterPro. The best match from each database was
94 assigned to the corresponding protein-coding gene. Following successful anchoring of 32
95 chromosomes and phasing of four sub-genomes by Hi-C (Figure S4A), we mapped the location of
96 upstream genes within the known biosynthetic gene cluster *HpBGC1*[20]. *HpBGC1* is located on Chr4A
97 within a topologically associating domain (TAD)[33] (Figure S4B), highlighting the co-regulation of the
98 cluster's genes for proper biosynthesis of the polyketide precursor PIBP. Intraspecies syntenic
99 analysis revealed the presence of five allele counterparts of *HpBGC1* on Chr4A-D and on the non-
100 homologous chromosome Chr5A (Figure S5A). These copies of *HpBGC1a–e* are not identical, but
101 bear variation in metabolic-gene co-linearity, gene-copy numbers and tandem duplications (Figure
102 S5B). By contrast, *HpBGC1d* in Chr4D has undergone substantial gene losses and retained only
103 one predicted *PKS*. These *HpBGC1* polymorphisms are possibly the result of segmental re-
104 arrangements within the tetraploid genome.

105 **Genome mining and co-expression analysis to identify aromatic prenyltransferases**

106 Several instances are reported of three or more genes comprising the same biosynthetic pathway
107 are physically co-localized, thus genome mining can aid gene discovery[33-36]. Because the polyketide-
108 module biosynthesis genes clustered within *HpBGC1*[20], clustering may well also be the case for the
109 downstream genes[32,37,38]. Based on functional annotation, 405 genes were predicted to be UbiA-
110 family aromatic prenyltransferases. By mining the *H. perforatum* genome, we identified a cluster
111 comprising 11 annotated aromatic-prenyltransferase genes bearing high homology to known
112 enzymes acting in biosynthesis of prenylated xanthones[39]. A blueprint of prenylated-xanthone
113 metabolism in *Hypericum* species[40-42] indicates analogies with the proposed model of hyperforin
114 biosynthesis involving the recruitment of dedicated CoA ligases, type-III polyketide synthases and
115 aromatic prenyltransferases that differ in the starting polyketide unit and the role of oxidative
116 cyclizations in the xanthone pathway[43,44] (Figure S6). Toward determining their functionality and
117 potential contribution to hyperforin biosynthesis in the absence of available functional-
118 characterization approaches in *H. perforatum*, the 11 predicted prenyltransferase genes were
119 individually cloned and co-expressed in yeast together with the known CoA ligase (*HpCCL2*) and
120 polyketide synthase (*HpPKS2*) from *HpBGC1*[20]. No prenylated products were detected by LC–MS
121 analysis, indicating that these prenyltransferases do not participate in hyperforin biosynthesis.

122 Genes in plant specialized-metabolite biosynthesis are usually co-regulated as part of the same
123 biological pathway, thus discovery of unknown metabolic genes in such pathways is usually based
124 on co-expression analysis of bulk RNA-seq data across plant samples (e.g. various tissues or
125 different treatments)[28,45-47]. Co-expression analysis with upstream genes from *HpBGC1* as bait did
126 not yield likely prenyltransferase candidates with high correlation efficiency[45,47] (Pearson Correlation
127 Coefficient > 0.90) among the 405 annotated prenyltransferases (Figure S7).



128

129 **Figure 1. Hyperforin biosynthesis, translucent cavities and genomic architecture of *Hypericum***
130 ***perforatum*.**

131 **A.** Overview of the proposed model of hyperforin biosynthesis *de novo* in St. John's wort. The pathway
132 involves three metabolic modules. The first module involves the biosynthesis of polyketide
133 phloroisobutyrophenone (PIBP) from valine[20]. The plastid-localized Methylerythritol 4-phosphate (MEP)

134 pathway biosynthesizes prenyl units like dimethylallyl-diphosphate (DMAPP) and geranyl-diphosphate
135 (GPP)[21]. The last module is hypothesized to involve prenylation of PIBP by a plastid-localized
136 transmembrane aromatic prenyltransferase(s) and an enigmatic cyclization involving the formation of a
137 C1–C8 bond[21-23].
138 **B.** Cross-section of *H. perforatum* leaf, highlighting the translucent pockets. Top: A tissue section stained
139 with tolonium chloride, with a schematic depiction (below) depicting the spatial profile of the indicated cell
140 types. Scale bar = 100 μm.
141 **C.** Overview of *H. perforatum* genome assembly. The CIRCOS plot represents the features of the *H.*
142 *perforatum* genome (size 1.39 Gb). From inner to outer layers: the gray links represent intraspecies
143 syntenic relationships, densities of simple sequence repeats, rRNA, tRNA, snRNA, miRNA, genes, GC
144 content, and transposable elements (TE) in 100-kb sliding windows, respectively. In the outer layer, bars
145 represent individual chromosomes. Blocks marked in blue on Chr4A (*HpBGC1*), in red on Chr5C (*HpPT3–*
146 *4*), and Chr8C (*HpPT1–2*) indicate metabolic gene locations involved in hyperforin biosynthesis. Blue
147 links highlight the syntenic relationships of *HpBGC1* copies among different chromosomes.

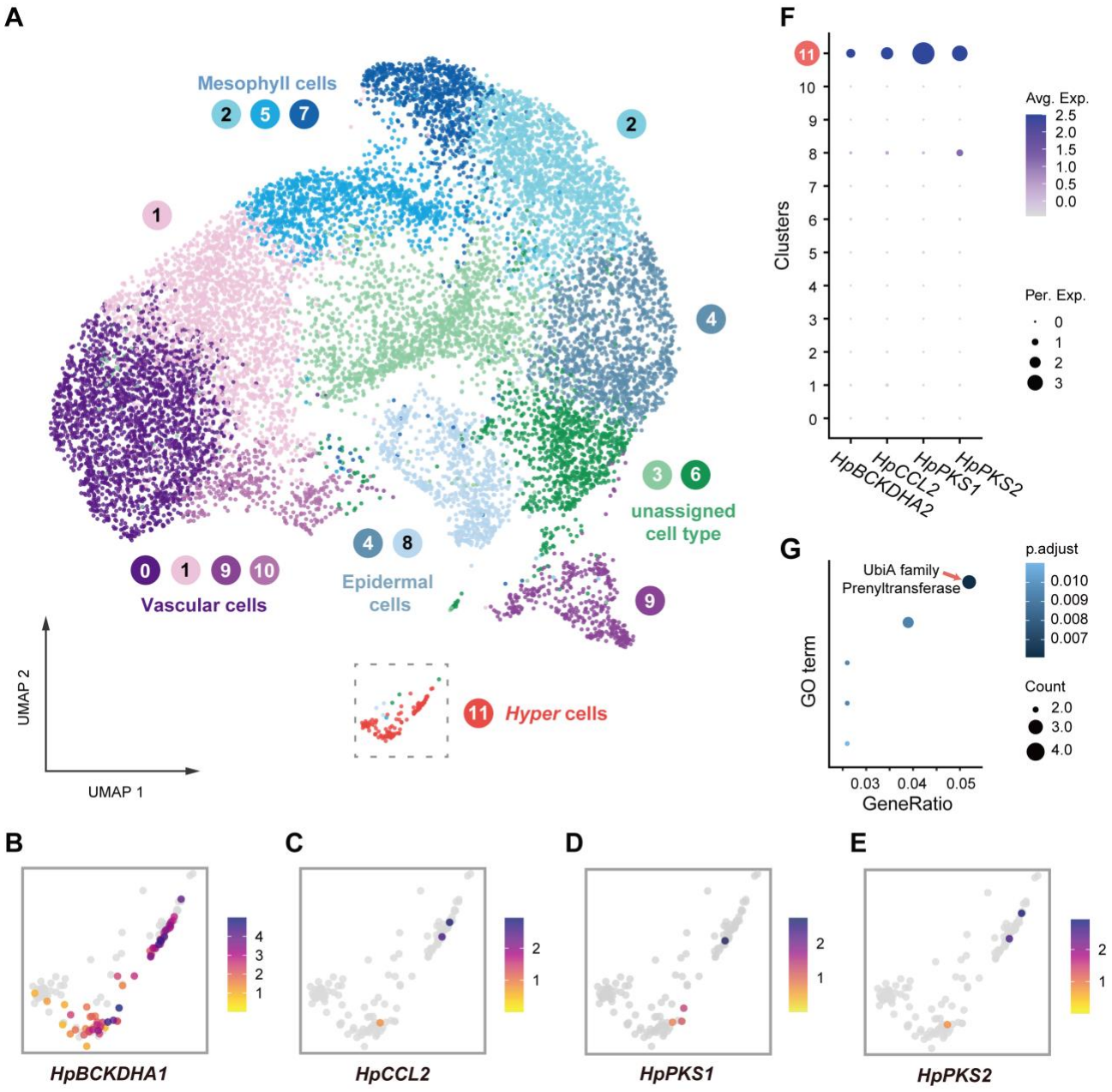
148 **Single-cell RNA sequencing enables discovery of missing hyperforin biosynthetic genes**

149 The eponym *perforatum* for St John's wort is derived from the visible translucent spots that
150 resemble perforations found on both leaves and flowers that also are present in certain other
151 *Hypericum* species[48,49] (Figure 1C). These translucent cavities are spherical cellular structures
152 spanning the upper to lower epidermis delimited by two cell layers — an inner layer consisting of
153 flattened secretory-like cells and an outer layer comprising turgid parenchymatous cells[50,51].
154 Hyperforin accumulates exclusively within these structures in *H. perforatum* leaves[51]. In contrast,
155 hyperforin is not detectable in non-secretory tissues, indicating that the hyperforin present in the
156 secretory structures originates exclusively from either the translucent cavities or from the
157 immediately adjacent cells[51]. It remains unclear whether the complete hyperforin biosynthesis, or
158 step/s in it, is active around the translucent cavities, or whether they simply serve as storage vessels.

159 High-throughput single-cell RNA sequencing (scRNA-seq) can distinguish cell types or cell
160 populations based on cell-to-cell heterogeneity as regards gene expression[52-55], but its mainstream
161 exploitation in plant science is hindered by substantial technical barriers such as the preparation of
162 protoplasts and the heterogeneity of plant tissues[52,54,55]. Despite this, scRNA-seq has been applied
163 to cellular differentiation and communication[52,54], while use in plant specialized metabolism is
164 sparse[53,56].

165 We hypothesize that since hyperforin biosynthesis correlates spatially with translucent cavities
166 and their associated cell types (Figure 1C), then the expression signal of hyperforin biosynthesis
167 genes might be concealed by noise typical of bulk RNA-seq, thus necessitating scRNA-seq to
168 carefully examine the cellular architecture of *H. perforatum* leaves. To generate a single-cell
169 transcriptome atlas of *H. perforatum* leaves, we successfully isolated viable protoplasts from young
170 leaves to capture the cellular heterogeneity within the tissue. Droplet-based 10x Genomics was used

171 for RNA-seq[57] to capture gene-expression data at the single-cell level. Reads were mapped to the
172 reference genome, quantifying gene-expression levels in individual cells with 92.2% of the reads
173 mapped to the reference genome, thus validating the quality and accuracy of the transcripome
174 sequence. We identified a total of 15,074 cells expressing a median of 1,405 genes across different
175 cell types. To interpret this dataset and categorize the cell types, low-quality and doublet cells were
176 removed, followed by normalization of the expression data to eliminate technical noise and
177 standardization for downstream analysis. Principal component analysis (PCA) was used to reduce
178 the dimensionality of scRNA-seq data and distill the complex gene-expression profiles, capturing the
179 most significant variation among the assayed cells. To categorize the cells into clusters based upon
180 shared transcriptome profiles, we applied the unbiased clustering algorithm uniform manifold
181 approximation and projection (UMAP)[58] to visualize and identify distinct cell populations (Figures 2A
182 and S8).



183

184 **Figure 2. A single-cell atlas of a *H. perforatum* leaf and identification of Hyper cells.**

185 **A**. Dimensionally reduced UMAP visualization of single leaf cells with each color representing a distinct
186 cluster. Twelve different cell clusters were identified. The four clusters of vascular cells (0, 1, 9 and 10)
187 are visualized in shades of purple–pink; the three clusters of mesophyll cells (2, 5 and 7) are visualized
188 in shades of blue–cyan; the two clusters of epidermal cells (4 and 8) in shades of blue–grey; and two
189 clusters of unassigned cells (3 and 6) in shades of green. Cells of cluster 11, visualized in red and
190 enclosed within a dashed box, are assigned as Hyper cells. Each dot represents one cell.
191 **B–E**. Close-up view of the expression profiles in Hyper cells of genes previously identified in hyperforin
192 biosynthesis *HpBCKDHA1*, *HpCCL2*, *HpPKS1* and *HpPKS2* from *HpBGC1*. The heatmap color scale
193 depicts the relative expression of each gene in each cell.
194 **F**. Dot plot illustrating the predominant expression of metabolic genes in the polyketide module of
195 hyperforin biosynthesis (*HpBCKDHA2*, *HpCCL2*, *HpPKS1* and *HpPKS2* from *HpBGC1*) across all
196 clusters in this atlas. The color scale indicates the average scaled expression of each gene in each cell
197 type. Dot sizes depict the fraction of cells wherein a given gene is expressed in a given cell type.
198 **G**. GO enrichment analysis of cluster 11, revealing 'UbiA-family prenyltransferase' as the top-enriched
199 protein family.

200 Annotation of cell types in a plant single-cell atlas is based on single or a set of marker genes.
201 However, such markers are absent for *H. perforatum*. There is, however, an extensive database of
202 genes from mainstream model plants used as cell-type markers[59-61]. Identification of functional
203 orthologs between *A. thaliana* and *H. perforatum* by sequence alignment and homology are prone
204 to false positives. Instead, synteny is one of the most reliable criteria for establishing the functional
205 relationships between orthologous genes or genomic regions across different species[62]. To annotate
206 the cell clusters with an unbiased method, we performed a fully automated annotation pipeline
207 comprising: (1) compiling an extensive marker-gene dataset for *A. thaliana*[59]; (2) through syntenic
208 analysis between *A. thaliana* and *H. perforatum*, we correlated the orthologous marker genes
209 between the two species and generated a marker-gene dataset for *H. perforatum*; and (3) using the
210 ScType[63] platform, an automated cell-type annotation was performed based on a pool of markers
211 (Figure S9). Twelve different cell types (Figures 2A and S8, Table S10) were identified with the cells
212 from clusters 0, 1, 9, 10 annotated as vascular cells, the cells from clusters 2, 5, 7 are mesophyll
213 cells, and cells from clusters 4 and 8 are epidermal cells. There were no adequate markers to support
214 annotation for cells from clusters 3 and 6, while the minority cell-type cluster 11 was annotated with
215 low confidence (i.e. low ScType score) as myrosin cells[59,64].
216 Visualization and quantification of expression of known early hyperforin biosynthetic-pathway
217 genes[20] revealed that *HpBCKDHA1*, *HpBCKDHA2*, *HpCCL2*, *HpPKS1* and *HpPKS2* are distinctively
218 and highly expressed in cluster 11 (Figure 2B–F). This indicates that biosynthesis of the precursor
219 PIBP takes place predominantly in cell-cluster 11, and in parallel, raises the question whether genes
220 encoding downstream steps are co-expressed in these cells. Toward addressing this, gene-ontology
221 (GO) enrichment analysis[65] found genes belonging to the UbiA prenyltransferase family are over-
222 represented in this cluster (Figures 2G and S10). The isoprenoid carbon atoms in hyperforin are

223 derived from the MEP plastidial pathway[21] and UbiA-type transmembrane aromatic
224 prenyltransferases are expected to catalyze the prenylation reactions by transferring isoprenoid units
225 to PIBP. MEP pathway genes *HPDXS*, *HpHDS1*, *HpHDS2*, *HpHDR* and *HpGPPS* are all highly
226 expressed in vascular cells (cluster 0) and in cluster 11 (Figure 3A,B). We therefore focused on
227 aromatic prenyltransferases expressed in cell-cluster 11. Based on this profile, four genes annotated
228 as aromatic prenyltransferases (*Hper_g146976*, *Hper_g147016*, *Hper_g96819* and *Hper_g96821*)
229 were selected as candidates based on their high expression in cluster 11 (Figure 3B,C), and
230 successfully cloned. All the selected candidate prenyltransferases have plastid transit peptides at
231 the N-termini, which were removed for expression in yeast based on ChloroP predictions[66].
232 To screen the candidate prenyltransferases involved in hyperforin biosynthesis (Figure 3D, E), we
233 assembled the PIBP precursor module in *Saccharomyces cerevisiae*. The CoA ligase HpCCL2 and
234 the type-III PKS HpPKS2[20] were cloned into pESC-Leu and transcripts lacking the corresponding
235 sequence encoding plastidial transit peptides of the four candidate prenyltransferases were cloned
236 into pESC-His and transformed into the isoprenoid-rich yeast strain DD104[67]. After galactose
237 induction, the culture was extracted with ethyl acetate and analyzed by LC–MS. Only the
238 prenyltransferase encoded by *Hper_g146976* demonstrated activity towards PIBP with a new peak
239 eluting at 9.9 min (Figures 3F and S11) under the extracted ion chromatogram (EIC) mode at *m*/z
240 333.2060, suggesting the enzyme uses a geranyl-diphosphate molecule as prenyl donor and
241 attaches the geranyl chain to PIBP. The geranyl-PIBP synthesized standard elutes at 9.9 min with
242 *m/z* = 330.2060 in positive mode, which aligned with the peak of the Hper_g146976 product (Figure
243 3F) and is in agreement with the MS^2 spectrum (Figures 3G and S11). We conclude that the
244 prenyltransferase encoded by *Hper_g146976* acts on PIBP and catalyzes the first prenylation step
245 in hyperforin biosynthesis by way of a geranyl transferase. It is therefore assigned HpPT1 and its
246 product as translucin I. Further analysis of enzyme-assay LC–MS data showed that HpPT1 uses
247 only GPP as a prenyl donor and not DMAPP (Figure S12).

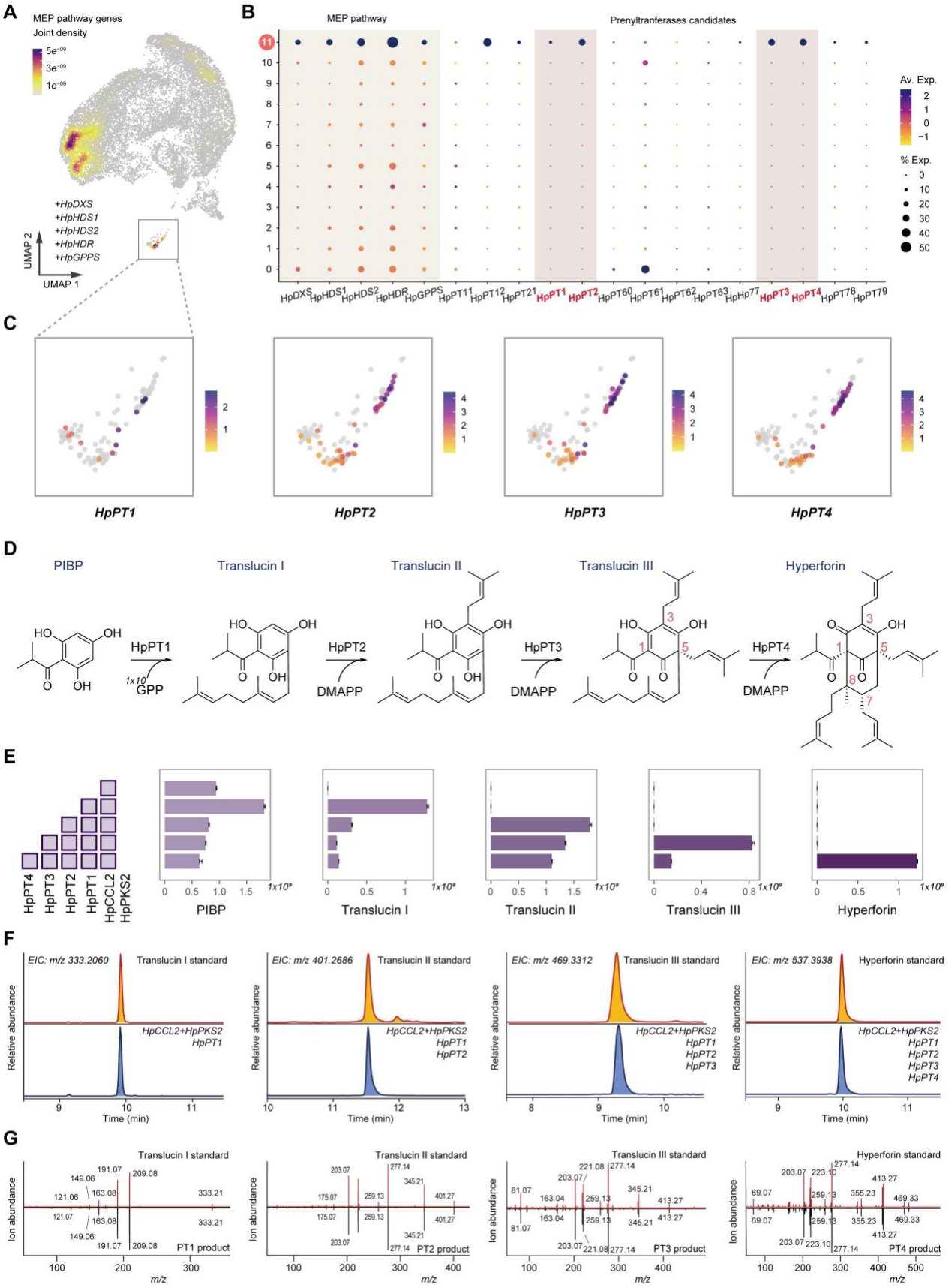


Figure 3. Elucidation of the catalytic steps to complete the formation of hyperforin.

A. Overlaid gene expression of MEP pathway genes (*HpDXS*, *HpHDS1*, *HpHDS2*, *HpHDR* and *HpGPPS*) across the *H. perforatum* leaf single-cell atlas demonstrates high expression in cells clusters 0 (vascular) and 11 (Hyper cells). The heatmap color scale indicates the sum of expression from the 5 MEP genes in each cell.

B. Dot plot highlighting elevated expression levels of MEP pathway genes in cluster 11, along with potential prenyltransferase (*PT*) candidate genes. Genes with confirmed activity are marked in red. The color scale indicates the average scaled expression of each gene in each cell type. Dot sizes reflect the fraction of cells wherein a given gene is expressed in a given cell type.

C. Cell-level expression profiles of prenyltransferases (*HpPT1*, *HpPT2*, *HpPT3* and *HpPT4*) in Hyper cells (cluster 11). The heatmap color scale shows the expression of each gene in each cell.

D. The sequential prenylation and tandem cyclization steps leading to hyperforin formation from phloroisobutyrophenone (PIBP), based on the verified activity of prenyltransferases HpPT1, HpPT2, HpPT3 and HpPT4. The intermediates translucins I–III were chemically synthesized and detected in *H. perforatum* extracts (Figures S11 and S13–15).

E. LC–MS peak area of the exact ion mass $[M + H]^+$ for each intermediate produced in yeast after co-expression of the indicated genes. Boxes to the left of the plots denote combinations of biosynthetic genes in each co-expression experiment. Data were obtained from three independent biological replicates and the bars represent the mean ± SE.

F. Comparison of extracted ion chromatogram (EIC) of each PIBP prenylated product from *in vivo* co-expression of biosynthetic genes (in red–gold) with EICs from corresponding synthetic standards (translucins I–III) and hyperforin (in black–blue).

G. Comparative MS/MS fragmentation patterns (MS^2 spectra) of prenylated products and standards from panel F**.**

To elucidate the outstanding biosynthesis steps involving PIBP prenylations, we cloned the three remaining candidate prenyltransferase genes into the second cloning site of pESC-His and assayed their activity in yeast. EtOAc extracts from galactose-induced strains were analyzed by LC–MS and a new peak at EIC *m/z* 401.2686 was detected at 11.6 min (Figures 3E,F and S13) in a strain expressing *Hper_g147016.* By comparison with the translucin-II synthetic standard and alignment of chromatograph (based on retention time, Figure 3F) and MS data (MS and MS^2, Figure 3G), we concluded that *Hper_g147016* encoding HpPT2 catalyzes the second prenylation in hyperforin biosynthesis.

To discover further the remaining prenylation steps, we cloned the remaining two candidates into pESC-Ura and assayed their activity using the same workflow. When we co-expressed *Hper_g96819* with HpCCL2, HpPKS2, HpPT1 and HpPT2, a new peak appeared in the LC–MS chromatogram of EIC *m/z* of 469.3312 at 9.4 min (Figures 3D–F and S14). This new product was identified as translucin III based on comparison of retention time, MS and MS^2 spectrum with a synthetic standard (Figures 3D–F and S14), indicating the geminal prenylation on C5 (Figure 3D) as

287 HpPT1 and the de-aromatization of the polyketide ring. Therefore, Hper_g96819 was assigned
288 HpPT3 that catalyzes the third prenylation step on the scaffold of PIBP polyketide.

289 Based on chemical intuition (Figure 3D), a prenylation and a cyclization step are expected for the
290 transformation of translucin III to hyperforin. To determine whether the only remaining candidate
291 gene encodes such functionalities, *Hper_g96821* was cloned into the second site of pESC-Ura and
292 co-expressed with HpCCL2, HpPKS2 and HpPT1–3. Hper_g96821 is enzymatically active,
293 transforming translucin III to hyperforin, with the product eluting at 10.0 min (the same time as the
294 hyperforin standard), and the MS and MS^2 spectra are identical to those of hyperforin (Figures 3D–
295 G and S15). Hper_g96821 catalyzes the fourth sequential prenylation step and was thus named
296 HpPT4.

297 Although HpPT4 is a UbiA prenyltransferase, it does not apparently catalyze a typical Friedel–
298 Crafts alkylation[24,26] as an aromatic prenyltransferases are expected to, but instead acts as an
299 irregular prenyltransferase catalyzing the 1′-2 coupling (branching)[25,68] between two isoprenyl units
300 (Figure S1A). Moreover, it appears that HpPT4 does not only catalyze the transfer of a prenyl unit to
301 translucin III, but also catalyzes bond formation between C(1) and C(8), resulting in cyclization and
302 biosynthesis of the iconic bicyclo[3.3.1]nonane (Fig. S1A) scaffold of PPAPs[17,18]. Such cyclizations
303 generating complex polycyclic structures are hallmarks of natural-product biosynthesis (Figure
304 S1B)[69].

305 To ensure host enzyme activities do not bias the functional characterization of candidates[70], we
306 assayed the activity of full-length HpPT1–4 by transient expression *Nicotiana benthamiana* leaves
307 by co-expression with *HpBCKDHA1*, *HpCCL2* and *HpPKS2* followed by LC–MS analysis of leaf
308 methanol extracts. Hyperforin was detected *in planta* only when all the enzymes in the pathway were
309 expressed (Figure 4A). This supports the functional characterization in yeast and safeguards the
310 proposed role of HpPT4 as a hyperforin synthase to transfer a prenyl group and catalyze the
311 cyclization step to complete hyperforin biosynthesis.

312 To further validate the proposed catalytic capacity of HpPT4 acting as branching 1′-2
313 prenyltransferase and cyclase, we tested the catalytic activity of HpPT4 against the translucin-III
314 analog colupulone[67] (Figure S16). We co-expressed HpPT4 with HpCCL2, HpPKS2, HlPT1L and
315 HlPT2[67] in yeast. LC–MS analysis of extracts showed that HpPT4 could transform colupulone to
316 secohyperforin[71] (8.0 min; *m/z* $[M+H]^+$ 469.3312) in a similar manner as it transforms translucin III to
317 hyperforin (Figure S16). To screen for additional prenyltransferases that have the same activity with
318 HpPT1–4; we cloned, expressed and assayed HpPT12, HpPT21 and HpPT77 (Figures S11, S13–
319 15). No activity was observed.

320

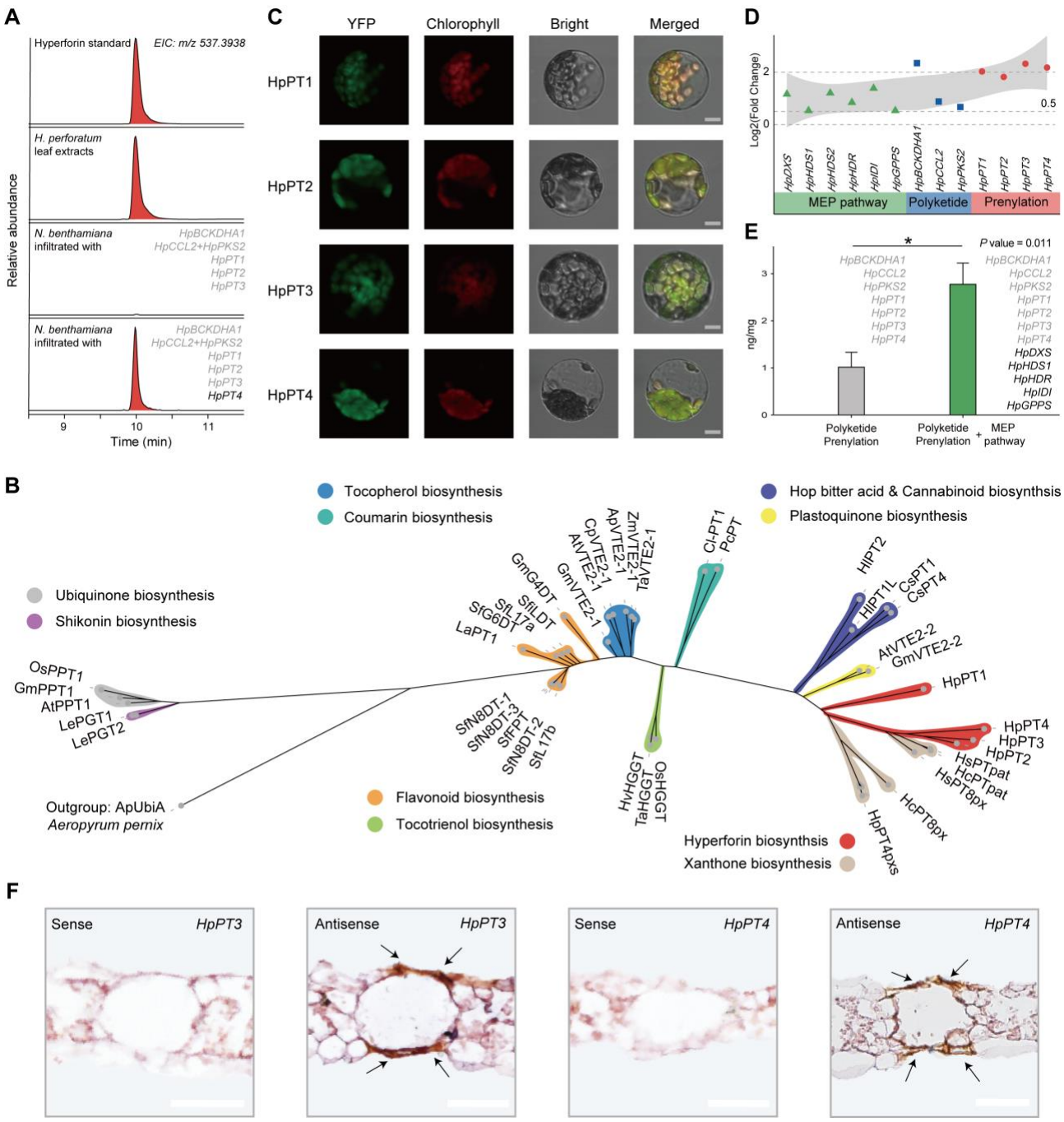


Figure 4: Plastids are prenyl-unit donors for hyperforin biosynthesis in *H. perforatum*.

A. Analysis of transient co-expression of *HpBCKDHA1*, *HpCCL2*, *HpPKS2*, *HpPT1*, *HpPT2*, *HpPT3* and *HpPT4* in tobacco for biosynthesis of hyperforin by LC–MS. Extracted ion chromatograms (EIC m/z 537.3938) from LC-MS analysis of a hyperforin standard and *H. perforatum* extract are shown for comparison.

B. Phylogenetic tree of characterized aromatic prenyl transferases from UbiA family from plants. The phylogeny was built from published, confirmed-active UbiA enzymes in plant metabolism and the prenyltransferases characterized in this study (highlighted in red shading) using 1,000 bootstrap values. UbiA from the fungus *Aeropyrum pernix* was used as the outgroup. The different color shades indicate the different pathways that the respective prenyltransferases are acting in. The phylogenetic analysis illustrating the close relation of prenyltransferases in hyperforin biosynthesis with the ones from plastid-localized plastoquinone biosynthesis (shaded in yellow).

334 **C**. Subcellular localizations of HpPT1, HpPT2, HpPT3 and HpPT4 with predicted transit peptides fused
335 at the C-terminus to YFP in transiently transgenic Arabidopsis leaf mesophyll protoplasts imaged by
336 confocal laser-scanning microscopy. Scale bars = 10 μm.
337 **D**. Dot plot depicting the logarithmic differential enhanced expression of genes in leaves from the
338 hyperforin biosynthetic modules MEP (green), polyketide/PIBP (blue) and prenylations (red) after 24 h of
339 treatment with 400 μM methyl jasmonate. All the genes in hyperforin biosynthesis *de novo* show
340 enhancement of their expression by methyl jasmonate.
341 **E.** LC–MS quantification of hyperforin production by transient co-expression in tobacco (hyperforin (ng)
342 per leaf sample dry weight (mg)) of the genes from polyketide and prenylation modules (in gray) and co-
343 expression of the genes from MEP, polyketide and prenylation modules (in green). Data were obtained
344 from five independent biological replicates. The bars represent the mean ± SE.
345 **F.** RNA *in situ* hybridization reveals the localization of *HpPT3* and *HpPT4* expression in proximity to the
346 translucent cavities in *H. perforatum* leaf cross-sections. Scale bars = 50 μm.

347 **Hyperforin biosynthesis *de novo* involving all three modules takes place in Hyper cells**

348 *HpPT1* and *HpPT2* are located close by on Chr8 (Figure S17A) and HpPT3 and HpPT4 are also
349 located near one another on Chr5 (Figure S17B). Genomic proximity can imply that the two pairs
350 are results of tandem duplications, however, the low sequence identity between these two pairs
351 suggest this is not the case (Table S11). Phylogenetic analysis of annotated prenyltransferases
352 based on amino-acid sequences reveals that HpPT1–4 are members of the HG clade of
353 transmembrane plant aromatic prenyltransferases with plastid localization, similar to enzymes acting
354 in the plastoquinone pathway (Figure 4B). HpPT1–4 fused at the C-terminus to YFP when transiently
355 expressed in *Arabidopsis* protoplasts localize to plastids (Figure 4C), supporting conclusions from
356 stable-isotope supplementation experiments indicating isoprenyl units are derived from the plastidial
357 MEP pathway[21].

358 Hyperforin biosynthesis can be divided in three metabolic modules involving the biosynthesis of
359 the polyketide PIBP localized to the cytosol, the MEP pathway in plastids supplying prenyl donors
360 (DMAPP, GPP), and plastid-localized transmembrane prenylation (Figure S18). We have previously
361 shown that genes encoding PIBP biosynthetic enzymes are up-regulated 24 h after methyl
362 jasmonate treatment[20]. Genes in the MEP pathway and *HpPT1–4* are also responsive to methyl
363 jasmonate treatment, indicating co-regulation of the whole hyperforin biosynthetic pathway (Figures
364 4D and S19). To confirm the role of *H. perforatum* genes from the MEP pathway in supplying
365 precursors for hyperforin biosynthesis, we transiently co-expressed the *H. perforatum* MEP genes
366 (*HpDXS*, *HpDXR*, *HpHDS1*, *HpHDR*, *HpIDI* and *HpGPPS*) together with genes from the PIBP
367 module and *HpPT1–4* in tobacco, resulting in over two-fold hyperforin production compared to co-
368 expression of genes from polyketide and prenylation modules (Figure 4E).

369 Because all the newly identified genes involved in hyperforin biosynthesis are exclusively and
370 highly expressed in cell-cluster 11, we therefore name them 'Hyper cells'. Using Seurat[58], we

identified 15 high-confidence genes that can be used as markers for Hyper cells (Table S12 and Figure S20). Of these, *HpBCKDHA1*, *HpPT2*, *HpPT3* and *HpPT4* feature with expression at least 4-fold higher than other cell types. To visualize the position of Hyper cells in leaf cellular architecture, we performed RNA *in situ* hybridization using probes against *HpPT3* and *HpPT4*. *HpPT3* and *HpPT4* transcripts indicate Hyper cells are adjacent to translucent cavities (Figures 4F, S21 and S22).

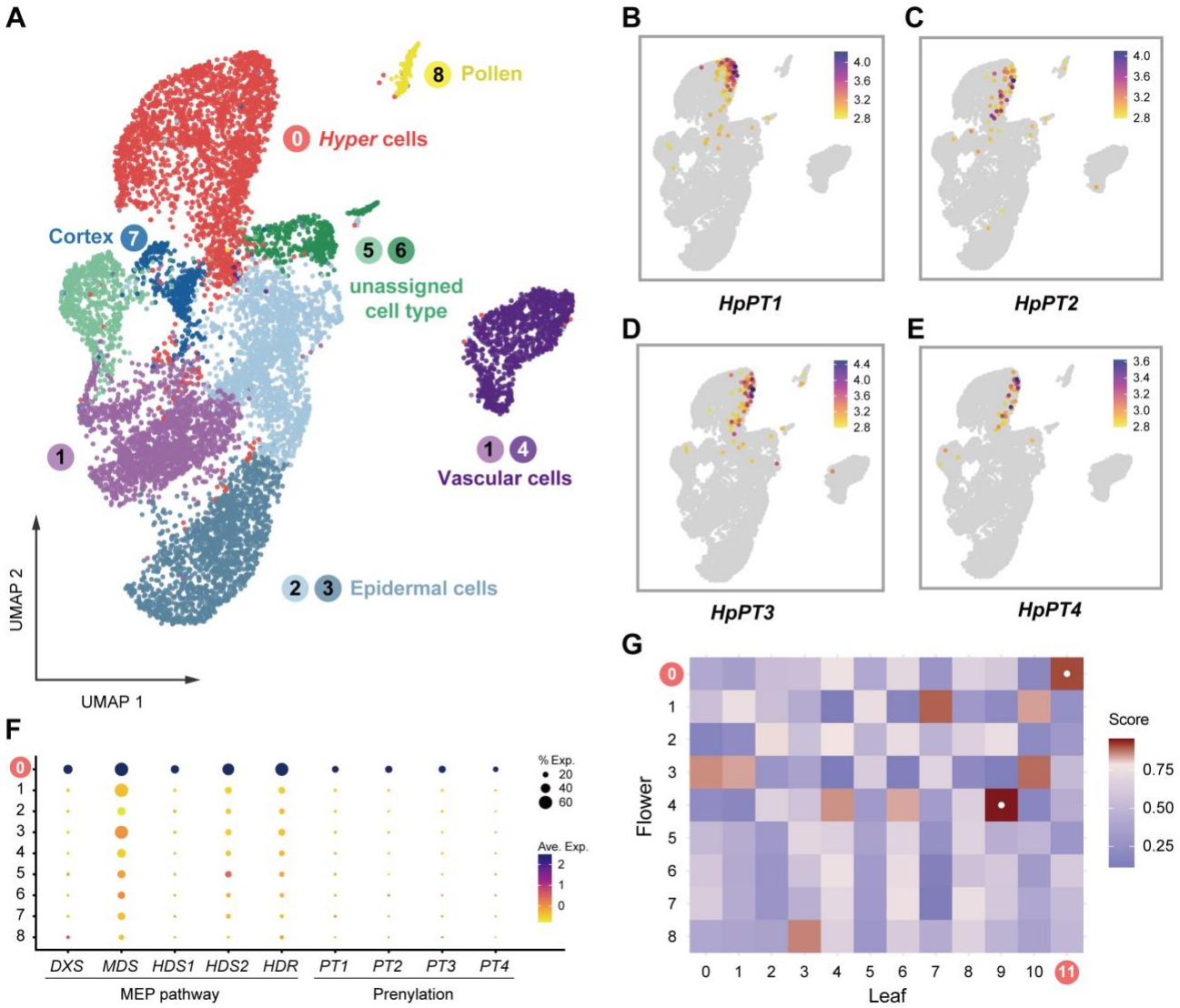


Figure 5: A single-cell atlas of *H. perforatum* flower reveals hyperforin biosynthesis in flowers.
A. Dimensionally reduced UMAP visualization depicting various single-cell clusters within the flower, with each color denoting a distinct cluster.
B–E. Expression profiles of the indicated prenyltransferases (PTs) at the single-cell level within the flower.
F. Dot plot showcasing the predominant expression patterns of metabolic genes involved in hyperforin biosynthesis *de novo* across identified clusters.
G. Similarity matrix comparing cell clusters between the single-cell datasets of the flower and leaf, highlighting correlations with scores above 0.90 with white dots.

Biosynthesis *de novo* of hyperforin in flowers also takes place in Hyper cells

386 Because hyperforin strongly accumulates in both leaves and open flowers[20,72], the discovery of
387 Hyper cells and elucidation of hyperforin biosynthesis *de novo* in leaves raises the question about
388 activity of the hyperforin biosynthetic pathway in flowers. To determine whether hyperforin
389 biosynthesis *de novo* in floral organs is organized at the cellular level in the same way as leaves,
390 protoplasts were isolated from flowers for scRNA-seq. The floral single-cell atlas reveals nine
391 different cell clusters (Figure 5A and Table S13). Based on a customized annotation pipeline, we
392 observed two types of epidermal cells (cell clusters 2 and 3), two types of vascular cells (cell clusters
393 1 and 4), one type of cells assigned as cortex cells (cell-cluster 7), one type of cells assigned as
394 pollen cells (cell-cluster 8), and two type of unassigned cells (cell clusters 5 and 6). *HpPT1–4* are
395 specifically and highly expressed in cell cluster 0 (Figure 5B–F) Moreover, expression of genes from
396 all three modules of hyperforin biosynthesis (MEP, PIBP/polyketide and prenylation) predominantly
397 occur in cell-cluster 0 in flowers (Figures 5B–F and S23).

398 To explore the correlation and conservation of different cell types in leaves and flowers, especially
399 regarding Hyper cells, we compared the respective scRNA-seq datasets using MetaNeighbor[73,74],
400 which through gene filtering, neighbor voting, and visualization, quantifies cell-type replicability,
401 determines gene sets that contribute to cell-type identity and enables rapid identification of clusters
402 with high similarity. A high level of similarity between cluster 11 from leaves with cluster 0 from flowers
403 lead to confirmation of Hyper cells being present in the two tissues of St. John's wort (Figure 5G).

404 **Discussion**

405 By sequencing the genome of *H. perforatum* at chromosome level and generating single-cell
406 atlases from leaves and flowers, we demonstrate how the combination of these approaches enabled
407 the elucidation of hyperforin biosynthesis *de novo* and reconstitution of the complete pathway in
408 yeast and tobacco. We anticipate these findings serve as a blueprint to elucidate the biosynthesis of
409 other prominent bioactive PPAPs such as the anticancer molecule garcinol[75] and neuroactive
410 garsubellin A[76] (Figure S1B). Our results illustrate how a UbiA-family aromatic prenyltransferase was
411 recruited to catalyze the 1'-2 irregular prenylation reaction[25] and also act as a terpene cyclase for
412 the formation of the bicyclo[3.3.1]nonane scaffold found in hyperforin and certain other PPAPs in
413 medicinal plants[17,18]. This complex bicyclic scaffold is a common structural feature among natural
414 products with different biosynthetic origins (such as alkaloids, terpenoids, polyketides and
415 meroterpenoids, Figure S1B)[46,77,78] across plants and microorganisms. Hyperforin synthase (HpPT4)
416 is one example of how nature's enzymatic versatility permits different enzymes to biosynthesize this
417 abundant yet complex bicyclic scaffold. Heterologous biosynthesis of hyperforin highlights potential
418 synthetic-biology applications for production of this valuable molecule and other bioactive PPAPs.

419 Single-cell RNA sequencing serves as a powerful, alternative approach for gene discovery and
420 rapid pathway elucidation in plant specialized metabolism, offering high resolution in gene
421 expression at the cellular level[53,56] . To address the challenge[55] of performing unbiased annotation of
422 the different cell types in non-model plant species as the majority of medicinal plants are, we

developed an automated annotation pipeline using ScType[63] with a customized marker-gene database based on identification of functional ortholog genes through syntenic analysis.

Our findings reveal a major role of Hyper cells — a previously unidentified type of cells — to function as cellular factories for the biosynthesis *de novo* of hyperforin that combines three biosynthetic modules (MEP, PIBP/polyketide and prenylation) in one cell type. The finding that Hyper cells are adjacent to translucent cavities where hyperforin is stored at high levels in leaves, together with the identification of biosynthetic genes, with our next steps toward identifying regulatory and transport mechanisms will deepen the understanding of specialized metabolism in plants at the single-cell level[79]. Our results not only provide a synthetic-biology approach to produce the hallmark polycyclic scaffold of hyperforin, but also underscore a powerful strategy for accelerating discovery and manipulation of natural-product biosynthesis in medicinal plants that are otherwise not readily amenable to functional characterization in conventional model organisms.

Materials and Methods

Plant materials

Hypericum perforatum subsp. *chinense* were collected in Enshi Tujia and Miao Autonomous Prefectures in Hubei, China, were delivered and grown under a 16 h:8 h light:dark cycle at 20°C in the CEMPS Phytotron.

Bulk RNA sequencing

A total of 17 plant tissues, namely untreated leaves, MeJA-treated leaves[20], young flowers, old flowers, young buds, middle-aged buds, old buds, leaves near buds, young leaves, middle-aged leaves, old leaves, stressed leaves, young stems, old stems, old roots, primary roots, and secondary roots were frozen on liquid nitrogen and homogenized using a tissue lyser (Shanghai Jingxin 40 JXFSTPRP-48). Total RNA was then extracted using the RNAprep Pure Plant Plus Kit (TIANGEN, Cat. #DP441), following the manufacturer's instructions.

The purity of the extracted RNA was assessed based on the $OD_{260\ nm}/OD_{280\ nm}$ ratio determined by a NanoDrop instrument (Thermo Fisher Scientific, USA) and agarose gel electrophoresis was performed to test for RNA degradation and potential contamination. RNA integrity was evaluated using a 2100 Bioanalyzer (Agilent Technologies, USA).

RNA samples that met quality-control criteria were used as input for preparation of Illumina short-read libraries. The NEBNext Ultra RNA Library Prep Kit (NEB, USA, Cat. #E7530L) was used according to the manufacturer's recommendations and index codes were incorporated into the libraries to attribute sequences to their respective samples. Qualified libraries were pooled together and subjected to sequencing on the Illumina HiSeq X Ten platform to produce 150-bp pair-end reads.

Mixed RNA from 11 different plant tissues (except untreated and MeJA-treated leaves) was prepared to construct continuous long-read (CLR) libraries using the SMRTbell Express Template Prep Kit 2.0 and sequenced on PacBio Sequel IIe to generate a full-length transcriptome.

Genome sequencing, assembly and annotation

For the long-read sequencing of the St. John's wort genome, genomic DNA was extracted using a phenol–chloroform method[80]. We constructed Single Molecule, Real-Time (SMRT) sequencing libraries following Pacific Biosciences standard protocols (www.pacb.com) by shearing DNA to a target size of 15–18 kb, followed by damage repair, end repair, blunt-end ligation and size selection to create a large insert SMRTbell library. This library was then sequenced on the PacBio Sequel platform. The assembly of the long reads was carried out using Hifiasm[81]. Additionally, a crosslinking step was employed using 4% v/v formaldehyde for the genomic DNA followed by vacuum treatment. After quenching the crosslinking reaction, samples were digested overnight with MboI. The fragment ends were marked with biotin and the chromatin DNA was re-ligated. The DNA underwent purification via the phenol–chloroform method, and biotin was removed from non-ligated fragment ends. A-tails were then added to the fragment ends and ligated to Illumina paired-end sequencing adapters. The resulting Hi-C sequencing libraries were amplified by PCR and sequenced on the Illumina NovaSeq platform. The data from Hi-C sequencing were used to anchor the long-read assembly into pseudochromosomes.

Repetitive genomic sequences were identified using TRF[82], RepeatMasker[83] and LTR_FINDER[84]. Structural annotation was performed using TblastN[85] and GeneWise[86] for homolog prediction and tools Augustus[87], Geneid[88], and Genescan[89] for *ab initio* gene prediction. Integration of RNA-seq data facilitated accurate gene-model prediction. Functional annotation involved aligning proteins to the Swiss-Prot database and domains to InterProScan. The prediction of non-coding RNAs, including tRNAs, rRNAs, miRNAs and snRNAs, was achieved using tRNAscan-SE[90] and alignment with the Rfam database. Genomic features were presented by different tracks in a CIRCOS plot[91].

Reference-genome-based read mapping

To acquire high-quality, clean RNA-seq reads, the raw data were trimmed using Trimmomatic v0.36[92] with the following parameters: HEADCROP: 13; LEADING: 3; TRAILING: 3; SLIDINGWINDOW: 4:15; and MINLEN: 36. Clean data were then aligned to the genome assembly using hisat2[93] with default settings. Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values were then generated through the cuffdiff command within the Cufflinks suite[94].

Preparation of St. John's Wort protoplasts for single-cell RNA-seq

Fresh St. John's Wort leaves and buds from 1 year old plants grown in the CEMPS Phytotron were collected and thoroughly washed with distilled water to remove any surface contaminants. Anthers were removed from bud samples because of their negative influence on protoplast purity, leaving only the petals and ovary. Tissues were then finely chopped and placed in a 10 cm culture dish. 10 mL of digestion solution (pH = 5.7) containing cellulase (1.25% w/v), pectinase (0.3% w/v), *D*-mannitol (0.4 M), MES (20 mM) and KCl (20 mM) was added to the culture dish containing the macerated tissues and incubated at room temperature (25 °C) with shaking at 80 rpm for 2 h. The

digested solution was centrifuged at room temperature (500 *g*) for 10 min to separate the protoplast-containing pellet from the tissue debris. The resulting protoplast pellet was carefully re-suspended in 500 µL washing buffer consisting of 0.4 mM mannitol, 20 mM MES pH 5.7, 20 mM KCl, 10 mM $CaCl_2$, and 0.1% (w/v) BSA. The protoplast suspension was filtered through a 40-µm cell sieve to remove any remaining tissue debris and obtain a clarified protoplast solution, followed by centrifugation at room temperature (200 *g*) for 6 min to further remove debris and obtain a purified protoplast pellet. The pellet was gently resuspended in 40 µL fresh washing buffer to obtain a concentrated protoplast suspension suitable for subsequent experiments. The viability of the isolated protoplasts was evaluated using trypan blue staining.

Single-cell (sc) RNA sequencing (scRNA-seq)

The library construction process used the Chromium Single Cell 3' Library & Single Cell 3' v3 Gel Beads kit. To generate single-cell GEMs (gel beads in emulsion), a suspension of protoplasts was loaded onto the automated Chromium Controller instrument. Ensuring the integrity of each GEM, reverse transcription occurred within them, resulting in cDNA that carried a unique cell barcode. Following cDNA synthesis, GEMs were broken to release the cDNA from individual cells. Liberated cDNA was then pooled and amplified using PCR. After amplification, the cDNA was fragmented and sequencing adapters were attached to construct libraries. Using the Illumina HiSeq X Ten platform, paired-end short reads of 150 bp were generated.

scRNA-seq data analysis

Raw sequencing data were preprocessed to transform it into a structured format by Cellranger[95]. Initial steps in Seurat (ref) involved filtering the dataset to exclude cells with aberrantly high gene counts (> 3,000), potentially indicative of artifacts or multiplets. This was followed by normalization to adjust for variations in sequencing depth across individual cells using 10,000 as scale factor. To focus the analysis on genes most likely to reveal biologically significant variation, a feature-selection process based on the top 2,000 variable genes was used. PCA dimensionality reduction was then applied to the data to distill the high-dimensional gene-expression profiles. The selection of dimensions for this reduction was determined through jackstraw and elbow plot[58]. The presence of potential doublets was assessed and removed by DoubletFinder[96]. Clustering algorithms were then used to identify distinct cell populations within the dataset. The resolution of clustering was fine-tuned by clustree[97] ranging from 0.1 to 1. For visual representation, the cell populations were projected by UMAP reduction. Marker genes were identified as genes predominantly expressed in each cell cluster. UCell[98] was used to score gene signatures based on the gene data matrix and MetaNeighbor[73,74] was used to quantify cell-type similarity between two datasets using neighbor voting.

We adapted a customized cell-type annotation approach using the ScType pipeline [63], due to the limitation of the standard ScType marker gene database being specific to human and mouse. We compiled an extensive marker gene dataset for *A. thaliana* from PlantscRNAdb [59]and mapped

these genes to their syntenic counterparts in *H. perforatum*. This tailored database enabled to perform automated and unbiased cell-type classification. A schematic of this annotation is detailed in Fig. S9.

Phylogenetic analysis of prenyltransferases

The amino-acid sequences of characterized prenyltransferases were aligned with sequences of genes obtained from synteny analysis using MUSCLE[99]. JTT+I+G was selected as the best-fit model of amino-acid replacement using prottest[100] based on alignment results. The maximum-likelihood method was then used to infer the phylogenetic tree by RaxML-ng[101] with 1,000 bootstrap values.

Gene isolation and plasmid construction

Total RNA was extracted from the aerial parts of *H. perforatum* and 5 μg was transcribed into cDNA using the SuperScriptTM IV First-Strand Synthesis system (ThermoFisher, Invitrogen). Candidate gene ORFs were PCR-amplified (Table S17) and cloned into various pESC vectors (pESC-LEU for *HpCCL2* and *HpPKS2*, pESC-HIS for *HpPT1* and *HpPT2*, pESC-URA for *HpPT3* and *HpPT4*), and pEAQ-HT (for *N. benthamiana* expression), PA7-YFP (yellow fluorescent protein was fused to the C-terminus of prenyltransferase signal peptides) using the ClonExpress II One Step Cloning Kit (Vazyme). Restriction enzymes and overhangs are detailed in Table Y. Following transformation into *E. coli* DH5α (DE3), colonies were selected on antibiotic-supplemented LB agar. Positive colonies were confirmed through colony PCR using primers listed in Table S17 and plasmids were isolated from cultures grown overnight at 37°C for verification by sequencing.

Expression and enzyme activity of prenyltransferases in yeast

The *Saccharomyces cerevisiae* DD104 strain (engineered for isoprenoid production)[102] was transformed using lithium acetate to introduce combinations of *HpCCL2*, *HpPKS2* and *PT* candidate genes inserted into pESC vectors (Agilent). Three days after transformation, colonies were grown in 5 mL dextrose dropout medium for 2 d. Following this, cells were harvested and re-suspended in 15 mL of dropout medium containing 2% (w/v) d-galactose for 3 d. After induction, cultures were extracted using an equal volume of ethyl acetate (EtOAc). The organic layer obtained was separated and dried under reduced pressure. The dried extract was subsequently resuspended in 400 µL MeOH followed by filtration using a 0.22-µm pore-size syringe filter.

Transient expression in *Nicotiana benthamiana*

For *Agrobacterium*-mediated transient expression in *N. benthamiana*, pEAQ-HT plasmids (Plant Bioscience Limited, Norwich, UK)[103] harboring genes of interest were transformed into *Agrobacterium tumefaciens* GV3101 with selection. PCR-positive colonies were cultured overnight in 15 mL LB medium with relevant selection. Cells were centrifuged and re-suspended in MMA buffer (10 mM $MgCl_2$, 10 mM MES, 150 µM acetosyringone), adjusted to $OD_{600\ nm}$ = 0.2 and kept in the dark for 2 h. Leaves of 4–6 week-old *N. benthamiana* were infiltrated with the indicated combinations

of each suspension. After 3 d, plants were supplemented with 2mM valine solution, and 2 d later, leaves were collected, flash-frozen and stored at –80°C. Frozen leaves were then lyophilized and ground into a fine powder then extracted in 10 mL methanol. The extraction was concentrated under vacuum, re-suspended in 500 µL methanol, filtered through a 0.22-µm organic syringe filter and stored at –20°C.

LC–MS method for metabolite analysis

Metabolites from co-expression experiments in yeast and tobacco samples were subjected to analysis using a Q Exactive system, integrated with a Dionex Ultimate 3000 UHPLC system. Three microliters (3 µL) of each processed sample was injected into a UPLC (Kinetex 2.6 µm C18 100 Å, measuring 100 x 2.1 mm, Phenomenex). This was carried out at a flow rate of 0.35 mL per minute. The gradient with solvent A (2 mM NH_4Ac) and solvent B (MeCN) from 0 to 0.5 min with maintained 10% B, followed by a linear increase from 10% to 98% B between 0.5–11 min. This was sustained at 98% B from 11–13 min. Mass spectra acquisition was performed in the positive mode under the following parameters: a spray voltage of 3,500 V, a capillary temperature of 320°C, a sheath gas of 40 and an aux gas of 10.

In situ RNA hybridization

RNA *in situ* hybridization was conducted using an optimized protocol[104]. Young *H. perforatum* leaves were fixed in FAA solution (3.7% v/v paraformaldehyde, 5% v/v acetic acid and 50% v/v ethanol), dehydrated by Leica HistoCore PEARL, and embedded in paraffin by Leica HistoCore Arcadia H. Thin sections (8 µm) were prepared using the HistoCore MULTICUT. The sections underwent toluidine blue staining for morphological examination. Additionally, a schematic representation of the leaf was generated by ggplantmap[105] for detailed visualization. Sense and antisense probes were synthesized through T_7-promoter-based *in vitro* transcription, incorporating digoxigenin-labeled UTP for subsequent detection. Hybridization was performed overnight at room temperature allowing for optimal probe–target annealing. Post-hybridization washes were conducted to remove excess probe using a series of saline-sodium citrate (SSC) buffers of decreasing concentration. The hybridized probes were then detected using an anti-digoxigenin antibody conjugated to alkaline phosphatase, with NBT/BCIP as the chromogenic substrate. The specificity of the hybridization signal was confirmed by comparing the results with sense probe controls. The hybridized sections were visualized using an Olympus BX53 microscope.

Subcellular localization of prenyltransferases

Arabidopsis thaliana protoplasts were prepared from Columbia-0 ecotype and transformed with PA7-YFP plasmids as described previously[20]. Yellow fluorescent protein (YFP) was fused to the C-terminus of predicted prenyltransferase signal peptides alone, giving rise to HpPT1–YFP (comprising amino acids 1–76), HpPT2–YFP (1–75), HpPT3–YFP (1–71) and HpPT4–YFP (1–79). The images were captured with a Zeiss LSM880 confocal microscope.

Chemical synthesis of prenylated intermediates

For translucin-I synthesis, PIBP (2.55 mmol) and geraniol (2.55 mmol) were dissolved in anhydrous CH_2Cl_2 (5 mL), followed by the addition of $BF_3 \cdot Et_2O$ (0.25 mmol) at 35°C. The reaction mixture was stirred for 1 h at the same temperature. Post-reaction, it was quenched with saturated $NaHCO_3$ solution (10 mL) and extracted with CH_2Cl_2 (3 × 20 mL). The dried organic phases (over Na_2SO_4) were then purified using silica gel column chromatography with EtOAc−hexane to yield translucin I[106].

For the synthesis of translucin II and III, translucin I (0.2 mmol) was dissolved in H_2O (2 mL) and THF (1 mL) under nitrogen stream at 0°C. KOH (0.6 mmol) was added, followed by prenyl bromide (0.8 mmol). After stirring for 1 h at 0°C, the reaction was quenched with saturated $NaHCO_3$ solution (10 mL) and extracted with CH_2Cl_2 (3 × 10 mL). The combined organic phases were dried over Na_2SO_4 and purified using silica gel column chromatography with EtOAc−hexane, resulting in translucin II and III[107].

Acknowledgments

We would like to thank the staff and management of the CEMPS Core Facility Center for the support in metabolomics (LC-MS, NMR) and microscopy services, as well as the personnel at the CEMPS Phytotron facilities. We also thank Prof. Bin Han and Dr Tao Huang for the access to computational servers for the bioinformatic analysis at the National Centre for Gene Research. We thank Wenjie Liang and Haojie Li for the help of microscope and Kexin Ji and Jiang Wang for the help of heterologous expression in tobacco. We acknowledge Prof. George Lomonossoff (John Innes Centre) for providing us with the pEAQ-HT plasmid. We acknowledge Prof. Guodong Wang (IGDB) for providing us with the DD104 yeast strain and the hops genes *HIPT1L*, *HIPT2*.

The work was financially supported by the National Natural Sciences Foundation of China, Research Fund for International Excellent Young Scientists (RFIS-II), grant 32150610477; Strategic Priority Research Program of the Chinese Academy of Sciences grant XDB27020204; Chinese Academy of Sciences, International Partnership Program of CAS grant 153D31KYSB20160074; and National Key Laboratory of Plant Molecular Genetics Special Fund. The CAS PIFI Fellowship and the China Postdoctoral Science Foundation for Postdoctoral International Exchange Program Fellowship provided fellowship support to Dr Ana Luisa Malaco Morotti.

636 **References**

637 1. Rizzo, P., Altschmied, L., Ravindran, B.M., Rutten, T., and D'Auria, J.C. (2020). The
638 Biochemical and Genetic Basis for the Biosynthesis of Bioactive Compounds in Hypericum
639 Perforatum L., One of the Largest Medicinal Crops in Europe. Genes (Basel) *11*.
640 10.3390/genes11101210.
641 2. Istikoglou, C., Mavreas, V., and Geroulanos, G. (2010). History and therapeutic properties of
642 Hypericum perforatum from antiquity until today. Psychiatriki *21*, 332-338.
643 3. Di Carlo, G., Borrelli, F., Ernst, E., and Izzo, A.A. (2001). St John's wort: Prozac from the plant
644 kingdom. Trends in Pharmacological Sciences *22*, 292-297.
645 4. Linde, K., Ramirez, G., Mulrow, C.D., Pauls, A., Weidenhammer, W., and Melchart, D. (1996).
646 St John's wort for depression—an overview and meta-analysis of randomised clinical trials.
647 Bmj *313*, 253-258.
648 5. Ng, Q.X., Venkatanarayanan, N., and Ho, C.Y.X. (2017). Clinical use of Hypericum
649 perforatum (St John's wort) in depression: A meta-analysis. Journal of Affective Disorders
650 *210*, 211-221. https://doi.org/10.1016/j.jad.2016.12.048.
651 6. Chatterjee, S.S., Bhattacharya, S.K., Wonnemann, M., Singer, A., and Müller, W.E. (1998).
652 Hyperforin as a possible antidepressant component of hypericum extracts. Life Sci *63*, 499-
653 510. 10.1016/s0024-3205(98)00299-9.
654 7. Schempp, C.M., Pelz, K., Wittmer, A., Schöpf, E., and Simon, J.C. (1999). Antibacterial
655 activity of hyperforin from St John's wort, against multiresistant Staphylococcus aureus and
656 gram-positive bacteria. Lancet *353*, 2129. 10.1016/s0140-6736(99)00214-7.
657 8. Leuner, K., Kazanski, V., Müller, M., Essin, K., Henke, B., Gollasch, M., Harteneck, C., and
658 Müller, W.E. (2007). Hyperforin--a key constituent of St. John's wort specifically activates
659 TRPC6 channels. Faseb j *21*, 4101-4111. 10.1096/fj.07-8110com.
660 9. Müller, W.E., Singer, A., and Wonnemann, M. (2001). Hyperforin--antidepressant activity by
661 a novel mechanism of action. Pharmacopsychiatry *34 Suppl 1*, S98-102. 10.1055/s-2001-
662 15512.
663 10. El Hamdaoui, Y., Zheng, F., Fritz, N., Ye, L., Tran, M.A., Schwickert, K., Schirmeister, T.,
664 Braeuning, A., Lichtenstein, D., Hellmich, U.A., et al. (2022). Analysis of hyperforin (St. John's
665 wort) action at TRPC6 channel leads to the development of a new class of antidepressant
666 drugs. Molecular Psychiatry *27*, 5070-5085. 10.1038/s41380-022-01804-3.
667 11. Robson, N.K.B. (2003). Hypericum Botany. In Hypericum: the genus Hypericum, E. Ernst, ed.
668 (CRC Press), pp. 1-22.
669 12. Barcaccia, G., Arzenton, F., Sharbel, T.F., Varotto, S., Parrini, P., and Lucchin, M. (2006).
670 Genetic diversity and reproductive biology in ecotypes of the facultative apomict Hypericum
671 perforatum L. Heredity *96*, 322-334. 10.1038/sj.hdy.6800808.
672 13. Nahrstedt, A., and Butterweck, V. (1997). Biologically active and other chemical constituents
673 of the herb of Hypericum perforatum L. Pharmacopsychiatry *30*, 129-134.

674 14. Tatsis, E.C., Boeren, S., Exarchou, V., Troganis, A.N., Vervoort, J., and Gerothanassis, I.P.
675 (2007). Identification of the major constituents of Hypericum perforatum by LC/SPE/NMR
676 and/or LC/MS. Phytochemistry *68*, 383-393. 10.1016/j.phytochem.2006.11.026.
677 15. Gurevich, A., Dobrynin, V., Kolosov, M., Popravko, S., and Riabova, I. (1971). Antibiotic
678 hyperforin from Hypericum perforatum L. Antibiotiki *16*, 510-513.
679 16. Richard, J.A. (2014). Chemistry and biology of the polycyclic polyprenylated
680 acylphloroglucinol hyperforin. European Journal of Organic Chemistry *2014*, 273-299.
681 17. Ciochina, R., and Grossman, R.B. (2006). Polycyclic polyprenylated acylphloroglucinols.
682 Chem Rev *106*, 3963-3986. 10.1021/cr0500582.
683 18. Yang, X.W., Grossman, R.B., and Xu, G. (2018). Research Progress of Polycyclic
684 Polyprenylated Acylphloroglucinols. Chem Rev *118*, 3508-3558.
685 10.1021/acs.chemrev.7b00551.
686 19. Ji, Y., Hong, B., Franzoni, I., Wang, M., Guan, W., Jia, H., and Li, H. (2022). Enantioselective
687 Total Synthesis of Hyperforin and Pyrohyperforin. Angewandte Chemie International Edition
688 *61*, e202116136. https://doi.org/10.1002/anie.202116136.
689 20. Wu, S., Malaco Morotti, A.L., Wang, S., Wang, Y., Xu, X., Chen, J., Wang, G., and Tatsis, E.C.
690 (2022). Convergent gene clusters underpin hyperforin biosynthesis in St John's wort. New
691 Phytologist *235*, 646-661. https://doi.org/10.1111/nph.18138.
692 21. Adam, P., Arigoni, D., Bacher, A., and Eisenreich, W. (2002). Biosynthesis of hyperforin in
693 Hypericum perforatum. J Med Chem *45*, 4786-4793. 10.1021/jm0209782.
694 22. Walsh, C.T., and Tang, Y. (2017). Natural product biosynthesis (Royal Society of Chemistry).
695 23. Dewick, P.M. (2011). Medicinal Natural Products: A Biosynthetic Approach (John Wiley &
696 Sons).
697 24. de Bruijn, W.J., Levisson, M., Beekwilder, J., van Berkel, W.J., and Vincken, J.-P. (2020).
698 Plant aromatic prenyltransferases: Tools for microbial cell factories. Trends in biotechnology
699 *38*, 917-934.
700 25. Christianson, D.W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. Chem
701 Rev *117*, 11570-11648. 10.1021/acs.chemrev.7b00287.
702 26. Tello, M., Kuzuyama, T., Heide, L., Noel, J., and Richard, S. (2008). The ABBA family of
703 aromatic prenyltransferases: broadening natural product diversity. Cellular and molecular life
704 sciences: CMLS *65*, 1459.
705 27. Vickery, C.R., La Clair, J.J., Burkart, M.D., and Noel, J.P. (2016). Harvesting the biosynthetic
706 machineries that cultivate a variety of indispensable plant natural products. Curr Opin Chem
707 Biol *31*, 66-73. 10.1016/j.cbpa.2016.01.008.
708 28. Reed, J., Orme, A., El-Demerdash, A., Owen, C., Martin, L.B.B., Misra, R.C., Kikuchi, S.,
709 Rejzek, M., Martin, A.C., Harkess, A., et al. (2023). Elucidation of the pathway for
710 biosynthesis of saponin adjuvants from the soapbark tree. Science *379*, 1252-1264.
711 10.1126/science.adf3727.

712 29. Zhao, Q., Yang, J., Cui, M.Y., Liu, J., Fang, Y., Yan, M., Qiu, W., Shang, H., Xu, Z., Yidiresi,
713 R., et al. (2019). The Reference Genome Sequence of Scutellaria baicalensis Provides
714 Insights into the Evolution of Wogonin Biosynthesis. Mol Plant *12*, 935-950.
715 10.1016/j.molp.2019.04.002.
716 30. Galla, G., Basso, A., Grisan, S., Bellucci, M., Pupilli, F., and Barcaccia, G. (2019). Ovule
717 Gene Expression Analysis in Sexual and Aposporous Apomictic Hypericum perforatum L.
718 (Hypericaceae) Accessions. Frontiers in Plant Science *10*. 10.3389/fpls.2019.00654.
719 31. Zhou, W., Wang, Y., Li, B., Petijová, L., Hu, S., Zhang, Q., Niu, J., Wang, D., Wang, S., Dong,
720 Y., et al. (2021). Whole-genome sequence data of Hypericum perforatum and functional
721 characterization of melatonin biosynthesis by N-acetylserotonin O-methyltransferase. J
722 Pineal Res *70*, e12709. 10.1111/jpi.12709.
723 32. Li, H., Wu, S., Lin, R., Xiao, Y., Malaco Morotti, A.L., Wang, Y., Galilee, M., Qin, H., Huang,
724 T., Zhao, Y., et al. (2023). The genomes of medicinal skullcaps reveal the polyphyletic origins
725 of clerodane diterpene biosynthesis in the family Lamiaceae. Mol Plant.
726 10.1016/j.molp.2023.01.006.
727 33. Nützmann, H.W., Doerr, D., Ramírez-Colmenero, A., Sotelo-Fonseca, J.E., Wegel, E., Di
728 Stefano, M., Wingett, S.W., Fraser, P., Hurst, L., Fernandez-Valverde, S.L., and Osbourn, A.
729 (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome
730 states. Proc Natl Acad Sci U S A *117*, 13800-13809. 10.1073/pnas.1920474117.
731 34. Polturak, G., Dippe, M., Stephenson, M.J., Chandra Misra, R., Owen, C., Ramirez-Gonzalez,
732 R.H., Haidoulis, J.F., Schoonbeek, H.-J., Chartrain, L., Borrill, P., et al. (2022). Pathogen-
733 induced biosynthetic pathways encode defense-related molecules in bread wheat.
734 Proceedings of the National Academy of Sciences *119*, e2123299119.
735 doi:10.1073/pnas.2123299119.
736 35. Méteignier, L.V., Nützmann, H.W., Papon, N., Osbourn, A., and Courdavault, V. (2023).
737 Emerging mechanistic insights into the regulation of specialized metabolism in plants. Nat
738 Plants *9*, 22-30. 10.1038/s41477-022-01288-7.
739 36. Polturak, G., and Osbourn, A. (2021). The emerging role of biosynthetic gene clusters in plant
740 defense and plant interactions. PLoS Pathog *17*, e1009698. 10.1371/journal.ppat.1009698.
741 37. Weng, J.K. (2014). The evolutionary paths towards complexity: a metabolic perspective. New
742 Phytol *201*, 1141-1149. 10.1111/nph.12416.
743 38. Leong, B.J., and Last, R.L. (2017). Promiscuity, impersonation and accommodation:
744 evolution of plant specialized metabolism. Current Opinion in Structural Biology *47*, 105-112.
745 39. Sayed, H.M.B., Nassar, S., Kaufholdt, D., Beerhues, L., Liu, B., and El-Awaad, I. (2023).
746 Biosynthesis of polyprenylated xanthones in Hypericum perforatum roots involves 4-
747 prenyltransferase. Plant Physiol *192*, 2971-2988. 10.1093/plphys/kiad219.
748 40. Fiesel, T., Gaid, M., Müller, A., Bartels, J., El-Awaad, I., Beuerle, T., Ernst, L., Behrends, S.,
749 and Beerhues, L. (2015). Molecular Cloning and Characterization of a Xanthone

750 Prenyltransferase from Hypericum calycinum Cell Cultures. Molecules *20*, 15616-15630.
751 10.3390/molecules200915616.
752 41. Nagia, M., Gaid, M., Biedermann, E., Fiesel, T., El-Awaad, I., Hänsch, R., Wittstock, U., and
753 Beerhues, L. (2019). Sequential regiospecific gem-diprenylation of tetrahydroxyxanthone by
754 prenyltransferases from Hypericum sp. New Phytol *222*, 318-334. 10.1111/nph.15611.
755 42. Sayed, H.M.B., Nassar, S., Kaufholdt, D., Beerhues, L., Liu, B., and El-Awaad, I. (2023).
756 Biosynthesis of polyprenylated xanthones in Hypericum perforatum roots involves 4-
757 prenyltransferase. Plant Physiology *192*, 2971-2988. 10.1093/plphys/kiad219.
758 43. Wang, Y., Malaco Morotti, A.L., Xiao, Y., Wang, Z., Wu, S., Chen, J., and Tatsis, E.C. (2022).
759 Decoding the Cytochrome P450 Catalytic Activity in Divergence of Benzophenone and
760 Xanthone Biosynthetic Pathways. ACS Catalysis *12*, 13630-13637.
761 10.1021/acscatal.2c03405.
762 44. El-Awaad, I., Bocola, M., Beuerle, T., Liu, B., and Beerhues, L. (2016). Bifunctional CYP81AA
763 proteins catalyse identical hydroxylations but alternative regioselective phenol couplings in
764 plant xanthone biosynthesis. Nat Commun *7*, 11472. 10.1038/ncomms11472.
765 45. Nett, R.S., Lau, W., and Sattely, E.S. (2020). Discovery and engineering of colchicine alkaloid
766 biosynthesis. Nature *584*, 148-153. 10.1038/s41586-020-2546-8.
767 46. Nett, R.S., Dho, Y., Tsai, C., Passow, D., Martinez Grundman, J., Low, Y.-Y., and Sattely, E.S.
768 (2023). Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis. Nature
769 *624*, 182-191. 10.1038/s41586-023-06716-y.
770 47. De La Peña, R., Hodgson, H., Liu, J.C.-T., Stephenson, M.J., Martin, A.C., Owen, C., Harkess,
771 A., Leebens-Mack, J., Jimenez, L.E., Osbourn, A., and Sattely, E.S. (2023). Complex scaffold
772 remodeling in plant triterpene biosynthesis. Science *379*, 361-368.
773 doi:10.1126/science.adf1017.
774 48. Maggi, F., Ferretti, G., Pocceschi, N., Menghini, L., and Ricciutelli, M. (2004). Morphological,
775 histochemical and phytochemical investigation of the genus Hypericum of the Central Italy.
776 Fitoterapia *75*, 702-711. https://doi.org/10.1016/j.fitote.2004.09.009.
777 49. Kucharíková, A., Kusari, S., Sezgin, S., Spiteller, M., and Čellárová, E. (2016). Occurrence
778 and Distribution of Phytochemicals in the Leaves of 17 In vitro Cultured Hypericum spp.
779 Adapted to Outdoor Conditions. Front Plant Sci *7*, 1616. 10.3389/fpls.2016.01616.
780 50. Ciccarelli, D., Andreucci, A.C., and Pagni, A.M. (2001). Translucent Glands and Secretory
781 Canals in Hypericum perforatum L. (Hypericaceae): Morphological, Anatomical and
782 Histochemical Studies During the Course of Ontogenesis. Annals of Botany *88*, 637-644.
783 10.1006/anbo.2001.1514.
784 51. Soelberg, J., Jørgensen, L.B., and Jäger, A.K. (2007). Hyperforin accumulates in the
785 translucent glands of Hypericum perforatum. Ann Bot *99*, 1097-1100. 10.1093/aob/mcm057.

786 52. Xu, X., and Jackson, D. (2023). Single-cell analysis opens a goldmine for plant functional
787 studies. Current Opinion in Biotechnology *79*, 102858.
788 https://doi.org/10.1016/j.copbio.2022.102858.
789 53. Li, C., Wood, J.C., Vu, A.H., Hamilton, J.P., Rodriguez Lopez, C.E., Payne, R.M.E., Serna
790 Guerrero, D.A., Gase, K., Yamamoto, K., Vaillancourt, B., et al. (2023). Single-cell multi-omics
791 in the medicinal plant Catharanthus roseus. Nature Chemical Biology *19*, 1031-1041.
792 10.1038/s41589-023-01327-0.
793 54. Liao, R.-Y., and Wang, J.-W. (2023). Analysis of meristems and plant regeneration at single-
794 cell resolution. Current Opinion in Plant Biology *74*, 102378.
795 https://doi.org/10.1016/j.pbi.2023.102378.
796 55. Grones, C., Eekhout, T., Shi, D., Neumann, M., Berg, L.S., Ke, Y., Shahan, R., Cox, K.L., Jr,
797 Gomez-Cano, F., Nelissen, H., et al. (2024). Best practices for the execution, analysis, and
798 data storage of plant single-cell/nucleus transcriptomics. The Plant Cell.
799 10.1093/plcell/koae003.
800 56. Kang, M., Choi, Y., Kim, H., and Kim, S.G. (2022). Single-cell RNA-sequencing of Nicotiana
801 attenuata corolla cells reveals the biosynthetic pathway of a floral scent. New Phytol *234*,
802 527-544. 10.1111/nph.17992.
803 57. Mazutis, L., Gilbert, J., Ung, W.L., Weitz, D.A., Griffiths, A.D., and Heyman, J.A. (2013).
804 Single-cell analysis and sorting using droplet-based microfluidics. Nature protocols *8*, 870-
805 891.
806 58. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J.,
807 Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell
808 data. Cell *184*, 3573-3587.e3529. 10.1016/j.cell.2021.04.048.
809 59. Chen, H., Yin, X., Guo, L., Yao, J., Ding, Y., Xu, X., Liu, L., Zhu, Q.H., Chu, Q., and Fan, L.
810 (2021). PlantscRNAdb: A database for plant single-cell RNA analysis. Mol Plant *14*, 855-857.
811 10.1016/j.molp.2021.05.002.
812 60. He, Z., Luo, Y., Zhou, X., Zhu, T., Lan, Y., and Chen, D. (2023). scPlantDB: a comprehensive
813 database for exploring cell types and markers of plant cell atlases. Nucleic Acids Research
814 *52*, D1629-D1638. 10.1093/nar/gkad706.
815 61. Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S., Wang, C., Xie, X., Gao, J., et al. (2021).
816 PCMDB: a curated and comprehensive resource of plant cell markers. Nucleic Acids
817 Research *50*, D1448-D1455. 10.1093/nar/gkab949.
818 62. Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny
819 and collinearity in plant genomes. Science *320*, 486-488. 10.1126/science.1153917.
820 63. Ianevski, A., Giri, A.K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type
821 identification using specific marker combinations from single-cell transcriptomic data. Nature
822 Communications *13*, 1246. 10.1038/s41467-022-28803-w.

823 64. Shirakawa, M., Ueda, H., Shimada, T., and Hara-Nishimura, I. (2016). FAMA: a molecular
824 link between stomata and myrosin cells. Trends in plant science *21*, 861-871.
825 65. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi,
826 M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res *51*,
827 D418-d427. 10.1093/nar/gkac993.
828 66. Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based
829 method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci *8*, 978-
830 984. 10.1110/ps.8.5.978.
831 67. Li, H., Ban, Z., Qin, H., Ma, L., King, A.J., and Wang, G. (2015). A heteromeric membrane-
832 bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations
833 in the bitter acid pathway. Plant Physiol *167*, 650-659. 10.1104/pp.114.253682.
834 68. Thulasiram, H.V., Erickson, H.K., and Poulter, C.D. (2007). Chimeras of Two Isoprenoid
835 Synthases Catalyze All Four Coupling Reactions in Isoprenoid Biosynthesis. Science *316*,
836 73-76. doi:10.1126/science.1137786.
837 69. Dang, T.T., Franke, J., Carqueijeiro, I.S.T., Langley, C., Courdavault, V., and O'Connor, S.E.
838 (2018). Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization
839 modes. Nat Chem Biol *14*, 760-763. 10.1038/s41589-018-0078-4.
840 70. Czechowski, T., Forestier, E., Swamidatta, S.H., Gilday, A.D., Cording, A., Larson, T.R.,
841 Harvey, D., Li, Y., He, Z., King, A.J., et al. (2022). Gene discovery and virus-induced gene
842 silencing reveal branched pathways to major classes of bioactive diterpenoids in
843 Euphorbia peplus. Proceedings of the National Academy of Sciences *119*,
844 e2203890119. doi:10.1073/pnas.2203890119.
845 71. Charchoglyan, A., Abrahamyan, A., Fujii, I., Boubakir, Z., Gulder, T.A., Kutchan, T.M.,
846 Vardapetyan, H., Bringmann, G., Ebizuka, Y., and Beerhues, L. (2007). Differential
847 accumulation of hyperforin and secohyperforin in Hypericum perforatum tissue cultures.
848 Phytochemistry *68*, 2670-2677. 10.1016/j.phytochem.2007.06.004.
849 72. Tekel'ová, D., Repák, M., Zemková, E., and Tóth, J. (2000). Quantitative changes of
850 dianthrones, hyperforin and flavonoids content in the flower ontogenesis of Hypericum
851 perforatum. Planta medica *66*, 778-780.
852 73. Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2018). Characterizing the
853 replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor.
854 Nature Communications *9*, 884. 10.1038/s41467-018-03282-0.
855 74. Fischer, S., Crow, M., Harris, B.D., and Gillis, J. (2021). Scaling up reproducible research for
856 single-cell transcriptomics using MetaNeighbor. Nature Protocols *16*, 4031-4067.
857 10.1038/s41596-021-00575-5.
858 75. Oike, T., Ogiwara, H., Torikai, K., Nakano, T., Yokota, J., and Kohno, T. (2012). Garcinol, a
859 histone acetyltransferase inhibitor, radiosensitizes cancer cells by inhibiting non-homologous
860 end joining. Int J Radiat Oncol Biol Phys *84*, 815-821. 10.1016/j.ijrobp.2012.01.017.

861 76. Fukuyama, Y., Kuwayama, A., and Minami, H. (1997). Garsubellin A, a novel polyprenylated
862 phloroglucin derivative, increasing choline acetyltransferase (ChAT) activity in postnatal rat
863 septal neuron cultures. Chemical and pharmaceutical bulletin *45*, 947-949.
864 77. Shen, X., Ting, C.P., Xu, G., and Maimone, T.J. (2020). Programmable meroterpene
865 synthesis. Nature Communications *11*, 508. 10.1038/s41467-020-14354-5.
866 78. Matsuda, Y., and Abe, I. (2016). Biosynthesis of fungal meroterpenoids. Nat Prod Rep *33*,
867 26-53. 10.1039/c5np00090d.
868 79. Weng, J.-K., Lynch, J.H., Matos, J.O., and Dudareva, N. (2021). Adaptive mechanisms of
869 plant specialized metabolism connecting chemistry to function. Nature Chemical Biology *17*,
870 1037-1045.
871 80. Gautam, A. (2022). Phenol-Chloroform DNA Isolation Method. In DNA and RNA Isolation
872 Techniques for Non-Experts, (Springer International Publishing), pp. 33-39. 10.1007/978-3-
873 030-94230-4_3.
874 81. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de
875 novo assembly using phased assembly graphs with hifiasm. Nat Methods *18*, 170-175.
876 10.1038/s41592-020-01056-5.
877 82. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic
878 Acids Res *27*, 573-580. 10.1093/nar/27.2.573.
879 83. Saha, S., Bridges, S., Magbanua, Z.V., and Peterson, D.G. (2008). Empirical comparison of
880 ab initio repeat finding programs. Nucleic Acids Res *36*, 2284-2294. 10.1093/nar/gkn064.
881 84. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length
882 LTR retrotransposons. Nucleic Acids Res *35*, W265-268. 10.1093/nar/gkm286.
883 85. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local
884 alignment search tool. J Mol Biol *215*, 403-410. 10.1016/s0022-2836(05)80360-2.
885 86. Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*,
886 988-995. 10.1101/gr.1865504.
887 87. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and
888 syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*,
889 637-644. 10.1093/bioinformatics/btn013.
890 88. Alioto, T., Blanco, E., Parra, G., and Guigó, R. (2018). Using geneid to Identify Genes. Curr
891 Protoc Bioinformatics *64*, e56. 10.1002/cpbi.56.
892 89. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic
893 DNA. J Mol Biol *268*, 78-94. 10.1006/jmbi.1997.0951.
894 90. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection
895 and functional classification of transfer RNA genes. Nucleic Acids Res *49*, 9077-9096.
896 10.1093/nar/gkab688.

897 91. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and
898 Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res
899 *19*, 1639-1645. 10.1101/gr.092759.109.
900 92. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
901 sequence data. Bioinformatics *30*, 2114-2120. 10.1093/bioinformatics/btu170.
902 93. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome
903 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915.
904 10.1038/s41587-019-0201-4.
905 94. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg,
906 S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq
907 reveals unannotated transcripts and isoform switching during cell differentiation. Nat
908 Biotechnol *28*, 511-515. 10.1038/nbt.1621.
909 95. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B.,
910 Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional
911 profiling of single cells. Nat Commun *8*, 14049. 10.1038/ncomms14049.
912 96. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in
913 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst *8*, 329-
914 337.e324. 10.1016/j.cels.2019.03.003.
915 97. Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings
916 at multiple resolutions. Gigascience *7*. 10.1093/gigascience/giy083.
917 98. Andreatta, M., and Carmona, S.J. (2021). UCell: Robust and scalable single-cell gene
918 signature scoring. Comput Struct Biotechnol J *19*, 3796-3798. 10.1016/j.csbj.2021.06.043.
919 99. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
920 throughput. Nucleic Acids Res *32*, 1792-1797. 10.1093/nar/gkh340.
921 100. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of
922 best-fit models of protein evolution. Bioinformatics *27*, 1164-1165.
923 10.1093/bioinformatics/btr088.
924 101. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast,
925 scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics
926 *35*, 4453-4455. 10.1093/bioinformatics/btz305.
927 102. Fischer, M.J., Meyer, S., Claudel, P., Bergdoll, M., and Karst, F. (2011). Metabolic engineering
928 of monoterpene synthesis in yeast. Biotechnol Bioeng *108*, 1883-1892. 10.1002/bit.23129.
929 103. Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P. (2009). pEAQ: versatile expression
930 vectors for easy and quick transient expression of heterologous proteins in plants. Plant
931 Biotechnology Journal *7*, 682-693. https://doi.org/10.1111/j.1467-7652.2009.00434.x.
932 104. Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., Gao, J., Yang, W., Lan, L., Wang,
933 Y., et al. (2021). An SHR–SCR module specifies legume cortical cell fate to enable nodulation.
934 Nature *589*, 586-590. 10.1038/s41586-020-3016-z.

105. Jo, L., and Kajala, K. (2023). ggPlantmap: an R package for creation of informative and quantitative ggplot maps derived from plant images. bioRxiv, 2023.2011.2030.569429. 10.1101/2023.11.30.569429.

106. Dethe, D.H., and Nirpal, A.K. (2021). Enantiospecific Total Synthesis of (-)-Japonicol C. Org Lett *23*, 2648-2653. 10.1021/acs.orglett.1c00560.

107. George, J.H., Hesse, M.D., Baldwin, J.E., and Adlington, R.M. (2010). Biomimetic synthesis of polycyclic polyprenylated acylphloroglucinol natural products isolated from Hypericum papuanum. Org Lett *12*, 3532-3535. 10.1021/ol101380a.