Supplementary Protocol

RNA-seq data analysis and identification of tissue-specific genes and transcription factors

Publicly available RNA-seq data from the Jia et al. study (Jia *et al.*, 2023) was used to identify tissue-specific genes and transcription factors (TFs) in five distinct tissues: stem, root, capsule, leaf and petal (**Supplementary Table 1**). The quality of the raw RNA-seq data was assessed using FastQC v0.11.9. (Andrews, 2010) and adapters were trimmed (where applicable) using Trim Galore v0.6.7 (https://github.com/FelixKrueger/TrimGalore). The clean data was aligned against the *Papaver somniferum* reference genome (cv. HN1; accession number GWHAZPJ00000000.1) (Yang *et al.*, 2021) using HISAT2 v2.2.1 (Kim *et al.*, 2019). The mapped reads were sorted using Samtools v1.14 (Danecek *et al.*, 2021) and transcript per million (TPM) counts were generated using StringTie v2.1.3 (Pertea *et al.*, 2015). Principal component analysis (PCA) plots were generated in R v4.3.2 (R Core Team, 2013) using the log2-normalised TPM counts to identify any outliers and ensure that the biological replicates clustered according to tissue type. Tissue-specific genes and TFs were identified using the mean TPM counts of biological replicates for each tissue type and the tau (T) metric (Kryuchkova-Mostacci and Robinson-Rechavi, 2017); a gene was deemed to be tissue-specific if it had the highest mean TPM count in one tissue type relative to the others and had a corresponding $T > 0.8$. Heatmaps showing gene expression of tissue-specific genes were generated using the *Z*-score of log2-normalised mean TPM counts and the pheatmap package in R (Thimm *et al.*, 2004). TFs were annotated using iTAK v1.7a33 (Zheng *et al.*, 2016).

Weighted gene co-expression network analysis (WGCNA)

Since co-expression may involve co-regulation, co-expression modules were generated to identify potential regulators (TFs) of known BIA biosynthesis pathway genes, especially the ones that are involved in the (*S*)-reticuline, noscapine and morphine pathways (**Supplementary Table S3**). First, gene-level quantification was performed on the mapped reads from the Jia et al. (2023) dataset using featureCounts from the Subread v2.0.6 package (Liao *et al.*, 2013, 2014). Genes with low expression variability (*p*-value 0.05) across samples were then filtered out using varianceBasedfilter from the R package DCGL (Liu *et al.*, 2010). A total of 13,037 (out of 55,316) genes remained; these genes were assigned to co-expression modules using the R package WGCNA (soft threshold =18; minClusterSize=50) (Langfelder and Horvath, 2008,

2012). Modules with similar gene expression profiles were merged (cutHeight=0.2) resulting in a total of 13 modules.

Gene regulatory network analysis

WGCNA identified a large number of TFs that are co-expressed with known BIA biosynthesis pathway genes. However, it is unlikely that every TF identified is a potential regulator of BIA biosynthesis. In order to determine which of these TFs share a regulatory link with the genes of interest, gene regulatory networks (GRNs) were constructed using SCION v4.0 (default parameters; clustering=none) (Clark *et al.*, 2021). Since SCION requires a minimum of five samples/replicates to build a network and the Jia et al. (2023) dataset only had three stem samples, we also included the stem samples from the Guo et al. (2018) dataset (Guo *et al.*, 2018).

Given the highly tissue-specific nature of the BIA biosynthesis genes from the three selected pathways ((*S*)-reticuline, noscapine and morphine), two GRNs were generated: a stem-specific GRN and a root-specific GRN. The data used to build the stem-specific network was as follows: 1) row-normalised TPM values of genes expressed in stem tissue (SRA accessions: SRR15146387, SRR15146388, SRR15146389, SRR6782291, SRR6782298); 2) a list of target genes, including BIA biosynthesis genes, that are stem-specific ($T > 0.8$) and found in co-expression modules (**Supplementary Table 2**); 3) a list of TFs that are stem-specific ($T > 0.8$) and found in co-expression modules (**Supplementary Table S4**). BIA biosynthesis genes that did not meet the criteria for tissue-specificity or found in co-expression modules were also included in the list of targets (**Supplementary Table 4**). The above steps were repeated to build the root-specific GRN except that root-specific data (SRA accessions: SRR15146362, SRR15146363, SRR15146367, SRR15146384, SRR15146385, SRR15146386), including root-specific genes and TFs (**Supplementary Table 2**) was used as input instead. Four subnetworks were extracted from the two GRNs: a stem-specific morphine subnetwork, a root-specific morphine subnetwork, a root-specific noscapine subnetwork and a root-specific (*S*)-reticuline subnetwork. The subnetworks were visualised in Cytoscape v3.10.2 (Shannon *et al.*, 2003). For ease of visualisation, the subnetworks were filtered based on the following edge weight cut-offs (0.5 for the stem-specific subnetwork and 0.75 for the root-specific subnetworks).

References

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom.

Clark NM, Nolan TM, Wang P, Song G, Montes C, Valentine CT, Guo H, Sozzani R, Yin Y, Walley JW. 2021. Integrated omics networks reveal the temporal signaling events of brassinosteroid response in Arabidopsis. Nature Communications **12**, 5858.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM. 2021. Twelve years of SAMtools and BCFtools. Gigascience **10**, giab008.

Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, Teodor R, Lu Y, Bowser TA, Graham IA. 2018. The opium poppy genome and morphinan production. Science **362**, 343-347.

Jia Y, Xu Y, Wang B, Guo L, Guo M, Che X, Ye K. 2023. The tissue-specific chromatin accessibility landscape of Papaver somniferum. Frontiers in Genetics **14**, 1136736.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology **37**, 907-915.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. Briefings in Bioinformatics **18**, 205-214.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics **9**, 1-13.

Langfelder P, Horvath S. 2012. Fast R functions for robust correlations and hierarchical clustering. Journal of statistical software **46**.

Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research **41**, e108-e108.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**, 923-930.

Liu B-H, Yu H, Tu K, Li C, Li Y-X, Li Y-Y. 2010. DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. Bioinformatics **26**, 2637-2638.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology **33**, 290-295.

R Core Team R. 2013. R: A language and environment for statistical computing.

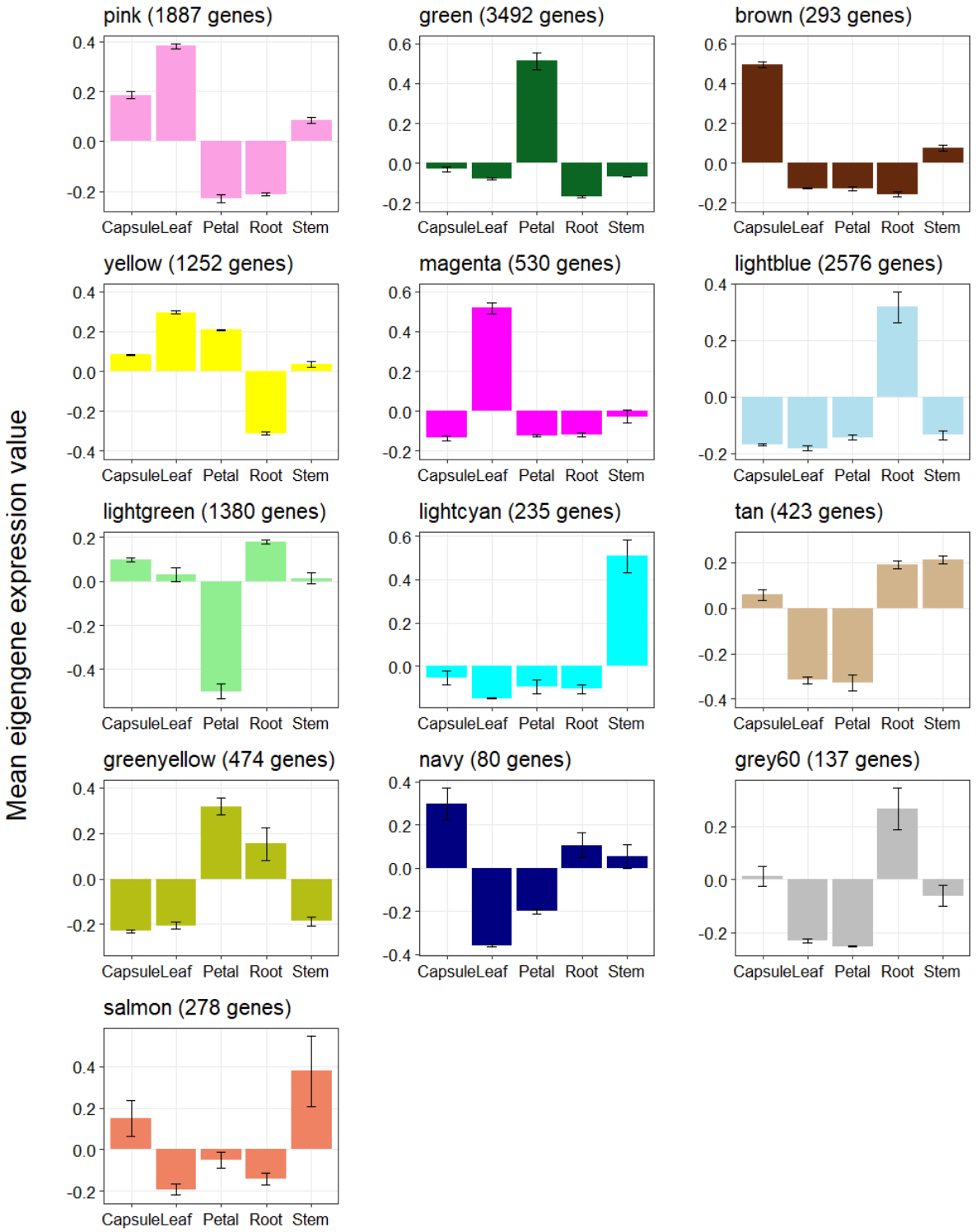
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research **13**, 2498-2504.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37**, 914-939.

Yang X, Gao S, Guo L, Wang B, Jia Y, Zhou J, Che Y, Jia P, Lin J, Xu T, et al. 2021. Three chromosome-scale Papaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nature Communications **12**, 6030.

Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ. 2016. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Molecular Plant **9**, 1667-1670.

Supplementary Figures



Supplementary Figure 1. Mean eigengene expression value for each tissue across the 13 co-expression modules identified by WGCNA. Numbers in parentheses represent the number of genes and transcription factors in each module. Values are mean ± standard error.